

Real-time Expressive Internet Communications

Zhe Xu

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

April 2005

Bournemouth University

Don. 2005

**BOURNEMOUTH
UNIVERSITY
LIBRARY**

M0013987 D0

03766098

SOR

004.678 XU

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and due acknowledgement must always be made to the use of any material contained in, or derived from this thesis.

Dedication

To XiaoYuan Xu because she is my best friend and my lovely wife

And

To Mom and Dad because they give me the opportunity to see the world

Real-time Expressive Internet Communications

Abstract: This research work – “Real-time Expressive Internet Communications” focuses on two subjects: One is the investigation of methods of automatic emotion detection and visualisation under real-time Internet communication environment, the other is the analysis of the influences of presenting visualised emotion – expressive images to Internet users.

To detect emotion within Internet communication, the emotion communication process over the Internet needs to be examined. An emotion momentum theory was developed to illustrate the emotion communication process over the Internet communication. It is argued in this theory that an Internet user is within a certain emotion state, the emotion state is changeable by internal and external stimulus (e.g. a received chat message) and time; stimulus duration and stimulus intensity are the major factors influencing the emotion state. The emotion momentum theory divides the emotions expressed in Internet communication into three dimensions: emotion category, intensity and duration.

The emotion momentum theory was implemented within a prototype emotion extraction engine. The emotion extraction engine can analyse input text in an Internet chat environment, detect and extract the emotion being communicated, and deliver the parameters to invoke an appropriate expressive image on screen to the every communicating user’s display. A set of experiments were carried out to test the speed and the accuracy of the emotion extraction engine. The results of the experiments demonstrated an acceptable performance of the emotion extraction engine.

The next step of this study was to design and implement an expressive image generator that generates expressive images from a single neutral facial image. Generated facial images are classified into six categories, and for each category, three different intensities were achieved. Users need to define only six control points and three control shapes to synthesise all the expressive images and a set of experiments were carried out to test the quality of the synthesised images. The experiment results demonstrated an acceptable recognition rate of the generated facial expression images.

With the emotion extraction engine and the expressive image generator, a test platform was created to evaluate the influences of emotion visualisation in the Internet communication

context. The results of a series of experiments demonstrated that emotion visualisation can enhance the users' perceived performance and their satisfaction with the interfaces.

The contributions to knowledge fall into four main areas; firstly, the emotion momentum theory that is proposed to illustrate the emotion communication process over the Internet; secondly, the innovations built into an emotion extraction engine, which senses emotional feelings from textual messages input by Internet users; thirdly, the innovations built into the expressive image generator, which synthesises facial expressions using a fast approach with a user friendly interface; and fourthly, the identification of the influence that the visualisation of emotion has on human computer interaction.

List of Publications resulting from this research work

- Boucouvalas, A. C, Xu, Z. and John, D., 2003. Expressive image generator for an emotion extraction engine. in: *Proceedings of the 17th HCI conference, Session 4, Track 2 Emotions and Computers*, University of Bath, UK, 367-381.
- Xu, Z. and Boucouvalas, A .C., 2002. Text-to-Emotion engine for real-time Internet communication. *In: Proceedings of the Second International Symposium on Communication Systems Networks and Digital Signal Processing (CSNDSP 2002)*, July 15-20, 2002, Staffordshire.
- Xu, Z., John, D. and Boucouvalas, A .C., 2002. *User preferences of a text-to-emotion engine In IEEE International Symposium on Consumer Electronics 2002 (ISCE 2002)*, Sept 23-26. 2002, Technical University Ilmenau, Germany.
- Xu, Z., John, D. and Boucouvalas,. A. C., 2003. Emotion extraction engine: expressive image generator. *in: Proceedings of EUROMEDIA*, Plymouth UK, 23-28.
- Xu, Z., John, D. and Boucouvalas, A. C., 2003. Social norm, expression and agent interaction. *submitted: Computers in Human Behavior*, 2003
- Xu, Z., John, D. and Boucouvalas, A. C., 2004. The influence of expressive images for computer interaction. *Accepted in Encyclopedia of Human Computer Interaction*.
- Xu, Z., John, D. and Boucouvalas, A. C., 2004. Internet interface design: social factors and design guidelines. *Accepted in Encyclopedia of Human Computer Interaction*.
- Xu, Z., John, D. and Boucouvalas,. A. C., 2004. Fuzzy logic usage in emotion communication of Human machine Interaction. *Accepted in Encyclopedia of Human Computer Interaction*.
- Xu, Z., John, D. and Boucouvalas, A. C., 2005. Expressive avatar generation: towards the expression automation. *Accepted in special issue of Journal of Visual Languages and Computing*.

Glossary

2D interface: Interfaces in which text and lines appear to be on the same flat level.

2.5D interface: The interfaces that apply various graphical algorithms to simulate the sense of depth on 2D interfaces.

3D interface: Interfaces in which text and images are not all on the same flat level.

ANOVA: Analysis of variance calculations used to test the hypothesis that means from two or more samples are equal.

CMC: Computer mediated communication. Humans using computers as the medium to communicate with other humans.

Cognitive style: The consistent underlying method of an individual's thinking and perceiving that affects the way in they perceive and respond to events and ideas.

Current emotion: Current emotion refers to the emotion contained in the most recent sentence.

Current mood: Current mood refers to the weighted average emotion in the five most recent sentences.

Degrees of Freedom: A value associated with the number of data points in a sample that is used in determining the significance level. The number of degrees of freedom varies depending on the type of calculation being performed.

Emotion: An excitement of the feelings caused by a specific exciting cause and manifested by some sensible effect on the body.

Emotional communication: The activity of communicating emotional feelings.

Emotion-evoking manner: The methods used to make readers perceive emotions.

Emotion decay: The gradual decline of the influence of emotion over time.

Emotion extraction engine: A software system which can extract emotions embedded in textual messages.

Emotion filter: Emotion filter is the mechanism used to detect and remove conflicting emotional feelings.

Emotion icon: A combination of keyboard characters or small images meant to represent a facial expression.

Eta Squared: The Magnitude of effect that assesses the degree to which variability among observations can be attributed to the tested variable.

Fuzzy logic: Fuzzy logic is applied to fuzzy sets where membership in a fuzzy set is a probability, not necessarily 0 or 1.

Personal Involvement Inventory: A measurement questionnaire developed by Zaichkowsky.

Power: The probability of a type II error (incorrectly concluding that there was no effect). The larger the observed power the more powerful the test (the maximum value that the power can be is 1.00). The smaller the power the more possibility that there is an effect may be missed.

Significance: An indicator that determines whether hypothesis should be accepted or rejected e.g. in ANOVA or t-test calculations that test the hypothesis that means from two or more samples are equal. A significance value of less than 0.05 ($p < 0.05$) indicates that the hypothesis is correct. A significance value of between 0.05 and 0.10 indicates that there is marginal evidence that the hypothesis is correct. A significance value of over 0.10 indicates that there is no evidence that the hypothesis is correct.

Software agent: A computer program that carries out tasks on behalf of another entity.

Sociability: The relative tendency or disposition to be sociable or associate with one's fellows

Social presence: The extent to which 'a person is perceived as a real person' in computer mediated communication

Task analysis: A method of providing an extraction of the tasks users undertake when interacting with system.

Wear down effect: A reduction in the participant's favourable responses after repeated exposures to a message.

Contents

REAL-TIME EXPRESSIVE INTERNET COMMUNICATIONS	V
LIST OF PUBLICATIONS RESULTING FROM THIS RESEARCH WORK.....	VII
GLOSSARY	VIII
LIST OF FIGURES.....	XVII
LIST OF TABLES.....	XX
PREFACE.....	XXII
ACKNOWLEDGEMENTS.....	XXII
CHAPTER 1 INTRODUCTION.....	1
1.1 MOTIVATION	2
1.2 STATEMENT OF THE PROBLEM	2
1.3 METHODOLOGY OF THIS STUDY	3
1.4 OUTLINE OF THIS THESIS	4
1.5 CONTRIBUTIONS TO THE KNOWLEDGE OF THE RESEARCH WORK.....	7
1.6 LIMITATIONS OF THE STUDY	8
1.7 CHAPTER CONCLUSION.....	9
CHAPTER 2 BACKGROUND REVIEW.....	10
2.1 EMOTION THEORIES	10
2.1.1 <i>The James-Lange theory</i>	10
2.1.2 <i>The Cannon-Bard theory</i>	11
2.1.3 <i>The facial-feedback theory</i>	11
2.1.4 <i>The Schachter-Singer theory</i>	12
2.1.5 <i>Cognitive appraisal theories</i>	13
2.2 CLASSIFICATIONS OF EMOTION	14
2.3 FUZZY LOGIC	15
2.3.1 <i>Fuzzy subsets</i>	15
2.3.2 <i>Fuzzy expert system</i>	15

2.4 EMOTION COMMUNICATION OVER THE INTERNET	16
2.4.1 <i>The importance of emotion communication over Internet interaction</i>	17
2.4.2 <i>Emotions in believable agents</i>	18
2.4.3 <i>Internet avatars</i>	18
2.5 COMPUTATIONAL EMOTIONS IN OTHER SCENARIOS	19
2.6 EMOTION DETECTION APPROACHES	20
2.6.1 <i>The self-report approach</i>	20
2.6.2 <i>The physiological approach</i>	21
2.6.3 <i>The facial expression recognition approach</i>	21
2.6.4 <i>Voice recognition approaches</i>	22
2.7 THE METHODS FOR TEXTUAL EMOTION EXTRACTION	22
2.7.1 <i>Rewrite-rule analysis</i>	23
2.7.2 <i>Keyword tagging</i>	25
2.7.3 <i>Lexical affinity</i>	25
2.7.4 <i>Statistical natural language processing</i>	25
2.7.5 <i>Real-world knowledge</i>	26
2.8 EMOTION VISUALISATION TECHNOLOGY – FACIAL EXPRESSION SYNTHESIS	26
2.8.1 <i>Image warping</i>	28
2.8.2 <i>Image morphing</i>	29
2.9 DESIGNING EASY-TO-USE INTERFACES	30
2.10 CHAPTER CONCLUSION	31
CHAPTER 3 THE PROPOSED THEORY AND PROTOTYPE SYSTEM OVERVIEW	32
3.1 THE EMOTION THEORY THIS STUDY FOLLOWED	32
3.1.1 <i>Cognitive appraisal theory</i>	32
3.1.2 <i>Ekman’s universal expression categories</i>	33
3.2 EMOTION MOMENTUM THEORY	34
3.3 THE PROPOSED SYSTEM ARCHITECTURE	36
3.4 CHAPTER CONCLUSION	39
CHAPTER 4 THE EMOTION EXTRACTION ENGINE	40
4.1 THE REQUIREMENTS OF THE EMOTION EXTRACTION ENGINE	40

4.1.1	<i>Interface requirements</i>	41
4.1.2	<i>Communications requirements</i>	41
4.1.3	<i>Input and output requirements</i>	42
4.1.4	<i>The requirements of the development language</i>	42
4.1.5	<i>Requirements of the performance assessment of the emotion extraction engine</i>	43
4.2	EMOTION EXTRACTION ENGINE SYSTEM ARCHITECTURE OVERVIEW	43
4.3	THE INTERFACE LAYER.....	46
4.4	ANALYSIS LAYER.....	47
4.4.1	<i>Input analysis</i>	48
4.4.2	<i>The tagging system</i>	49
4.4.2.1	Tagged dictionary.....	49
4.4.2.2	Tag set.....	50
4.4.2.3	Word analysis component.....	52
4.4.3	<i>The parser</i>	53
4.4.3.1	Individual sentence parsing.....	54
4.4.3.2	Mood analysis and conflicting emotion detection.....	58
4.4.4	<i>The limitations of the emotion extraction engine</i>	63
4.5	DEVELOPMENT SUMMARY	65
4.6	THE PERFORMANCE ASSESSMENT OF THE EMOTION EXTRACTION ENGINE	65
4.6.1	<i>Method of the accuracy experiment</i>	65
4.6.2	<i>The operation of the accuracy experiment</i>	66
4.6.3	<i>Results of the accuracy experiment</i>	67
4.6.4	<i>Accuracy experiment result analysis</i>	67
4.6.5	<i>Engine speed experiment operation</i>	68
4.6.6	<i>Engine speed experiment results and analysis</i>	69
4.7	CONFLICTING EMOTIONS EXPERIMENT	69
4.7.1	<i>Planning the experiment</i>	69
4.7.2	<i>The design of the conflicting emotion experiment</i>	70
4.7.2.1	Conflicting emotion story 1.....	70
4.7.2.2	Conflicting emotion story 2.....	71
4.7.2.3	Conflicting emotion story 3.....	72
4.7.3	<i>Conflict emotion experiment results and analysis</i>	73
4.8	CHAPTER CONCLUSION.....	73

CHAPTER 5 FUZZY EMOTION ANALYSIS FUNCTIONS	74
5.1 CONFLICTING EMOTION DETECTION COMPONENT	75
5.2 MOOD SELECTION COMPONENT	76
5.3 EMOTION STORAGE COMPONENT	85
5.4 EMOTION FILTER COMPONENT	85
5.5 DEVELOPMENT SUMMARY	89
5.6 CONFLICTING EMOTION EXPERIMENT	89
5.6.1 <i>Experiment planning and methods</i>	89
5.6.2 <i>Experiment result analysis</i>	90
5.7 CHAPTER CONCLUSIONS	90
CHAPTER 6 EXPRESSIVE IMAGE GENERATOR.....	92
6.1 THE REQUIREMENTS OF THE EXPRESSIVE IMAGE GENERATOR.....	93
6.2 THE DEVELOPMENT OF THE EXPRESSIVE IMAGE GENERATOR.....	95
6.2.1 <i>Expressive characteristics</i>	96
6.2.2 <i>Image warping subsystem</i>	97
6.2.2.1 Expression model mask definition	97
6.2.2.2 Start point and control area collection	98
6.2.2.3 Finish points definition.....	99
6.2.2.4 The transform function T.....	102
6.2.3 <i>Image morphing subsystem</i>	104
6.2.4 <i>Facial expression animation</i>	107
6.2.5 <i>Development summary</i>	109
6.3 EXPRESSIVE IMAGE GENERATOR EXPERIMENTS	110
6.3.1 <i>Planning the experiment</i>	110
6.3.2 <i>Method of the first experiment</i>	110
6.3.3 <i>Method of the second experiment</i>	112
6.3.4 <i>Experiment summary</i>	113
6.4 CHAPTER CONCLUSIONS	114
CHAPTER 7 APPLICATIONS OF THE EMOTION EXTRACTION ENGINE	116
7.1 STOCK MARKET EMOTION ANALYSER: TAILORING THE EMOTION EXTRACTION ENGINE.....	116

7.1.1 Stock market analysis articles	117
7.1.2 Emotion definition in the stock analyser	118
7.1.3 Emotion measurement.....	118
7.1.4 Stock market emotion analyser architecture	119
7.1.5 Stock market analyser experiment	121
7.1.5.1 Methods of the experiment.....	122
7.1.5.2 Percentage analysis	122
7.1.5.3 Correlation analysis.....	123
7.1.6 Stock market emotion analyser conclusions.....	127
7.2 CHAT ENVIRONMENT APPLICATIONS	127
7.3 CHAPTER CONCLUSION	129
CHAPTER 8 PREFERENCE EXPERIMENT ON EMOTION EXTRACTION ENGINE	131
8.1 COGNITIVE STYLE	131
8.1.1 Definitions of cognitive style.....	131
8.1.2 The Wholist-Analytic and Verbal-Imagery dimensions of cognitive style.....	131
8.1.3 The Cognitive Styles Analysis program	133
8.2 PLANNING THE EXPERIMENT.....	134
8.3 METHOD OF THE EXPERIMENTS.....	134
8.4 EXPERIMENT OPERATION.....	139
8.5 HOW PREFERENCE WAS MEASURED AND ANALYSED	139
8.6 EXPERIMENT RESULTS.....	140
8.6.1 Question analysis.....	140
8.6.2 Experiment results from CSA tests.....	144
8.6.2.1 Interface preferences between different cognitive style groups.....	145
8.6.2.2 Component preferences by cognitive style	145
8.7 CHAPTER CONCLUSIONS	146
CHAPTER 9 AN EXPERIMENT ON ASSESSING THE PREFERENCE DIFFERENCE BETWEEN 2D AND 3D EXPRESSIVE INTERFACES	148
9.1 BACKGROUND	148
9.1.1 Social presence.....	149
9.1.2 Human sociability style.....	149
9.1.3 Activity style.....	150

9.2 PLANNING THE EXPERIMENT.....	151
9.3 EXPERIMENT PLATFORM DEVELOPMENT.....	151
9.4 SOCIAL PRESENCE DIFFERENCES BETWEEN 2D AND 3D INTERFACES	152
9.5 METHOD OF THE EXPERIMENT	156
9.6 EXPERIMENT OPERATION.....	158
9.7 EXPERIMENT RESULTS.....	159
9.7.1 Percentage analysis.....	159
9.7.2 Activity style Vs interface preference.....	160
9.7.3 Sociability Vs interface preference.....	160
9.7.4 Gender Vs the interface preference.....	161
9.7.5 The analysis of the hypotheses	162
9.8 CHAPTER CONCLUSIONS.....	162
CHAPTER 10 THE INFLUENCE OF EXPRESSIVE IMAGES DISPLAY ON HCI AND THE FACTORS AFFECTING THE INFLUENCES.....	164
10.1 THE EXPERIMENT ASSESSING THE INFLUENCE OF EXPRESSIVE IMAGE DISPLAY	164
10.1.1 Background.....	164
10.1.1.1 Social norms.....	164
10.1.1.2 Emotions and social norms	165
10.1.1.3 Social norms and computer communication	166
10.1.1.4 Emotional agent.....	166
10.1.2 Planning the experiment	167
10.1.3 Methods of the experiment	168
10.1.3.1 The online quiz.....	168
10.1.3.2 Questionnaire.....	172
10.1.4 Experiment environment	172
10.1.5 How the data was measured and analysed.....	173
10.1.6 Experiment results	173
10.1.7 Experiment conclusions	181
10.2 FACTORS THAT MAY AFFECT THE INFLUENCE OF EXPRESSIVE IMAGES.....	182
10.2.1 Background.....	182
10.2.2 Planning of experiment.....	183
10.2.3 Methods of the experiment	184

10.2.3.1	The emotion intensity test.....	184
10.2.3.2	Display duration - wear down effect test	185
10.2.4	<i>The experiment operation.....</i>	<i>186</i>
10.2.4.1	The intensity test operation.....	186
10.2.4.2	The wear down test operation	186
10.2.5	<i>How the data was measured and analysed.....</i>	<i>186</i>
10.2.6	<i>Experiment result analysis</i>	<i>187</i>
10.2.6.1	Intensity test results.....	187
10.2.6.2	The wear down result analysis	187
10.2.7	<i>Experiment results discussion</i>	<i>187</i>
10.2.8	<i>Experiment conclusions</i>	<i>189</i>
10.3	CHAPTER CONCLUSION.....	189
 CHAPTER 11 CONCLUSIONS AND FURTHER RESEARCH SUGGESTIONS.....		190
11.1	CONCLUSIONS AND FURTHER RESEARCH SUGGESTIONS FOR THE EMOTION EXTRACTION ENGINE PROTOTYPE	191
11.2	CONCLUSIONS AND FURTHER RESEARCH SUGGESTIONS FOR THE EXPRESSIVE IMAGE GENERATOR	193
11.3	CONCLUSIONS AND RESEARCH SUGGESTIONS FOR EMBEDDING EMOTION EXTRACTION ENGINE INTO DIFFERENT APPLICATIONS	194
11.4	CONCLUSIONS AND FURTHER RESEARCH SUGGESTIONS FOR EXPERIMENTS ASSESSING THE INFLUENCE OF VISUALISED EMOTIONS	195
11.4.1	<i>The interface preference experiment.....</i>	<i>195</i>
11.4.2	<i>The experiment comparing the preference of the 2D and 3D interfaces.....</i>	<i>196</i>
11.4.3	<i>The experiment assessing the influences of emotion visualisation.....</i>	<i>196</i>
11.4.4	<i>The experiment assessing the factors that may affect perceived emotional feelings.....</i>	<i>197</i>
REFERENCES		198

List of figures

FIGURE 1: THE PROCESS OF THE JAMES-LANGE THEORY	11
FIGURE 2: THE PROCESS OF CANNON-BARD THEORY	11
FIGURE 3: THE PROCESS OF THE FACIAL EXPRESSION THEORY	12
FIGURE 4: THE PROCESS OF THE SCHACHTER-SINGER THEORY	13
FIGURE 5: COGNITIVE APPRAISAL THEORY	13
FIGURE 6: STOCK AND USER INTERACTION.....	20
FIGURE 7: TREE PRESENTATION FOR NOUNS	23
FIGURE 8: TREE PRESENTATION OF A SENTENCE	24
FIGURE 9: THE WARPING PSEUDO CODE.....	28
FIGURE 10: SEGMENT WARPING AREA.....	29
FIGURE 11: EMOTION MOMENTUM	34
FIGURE 12: THE WORKFLOW OF THE PROTOTYPE SYSTEM.....	37
FIGURE 13: OVERVIEW OF THE EMOTION EXTRACTION ENGINE.....	44
FIGURE 14: STRUCTURE OVERVIEW OF THE ANALYSIS LAYER.....	45
FIGURE 15: ARTICLE ANALYSIS INTERFACE	46
FIGURE 16: EXPRESSIVE 2D CHAT INTERFACE.....	47
FIGURE 17: THE INPUT ANALYSIS FUNCTION	49
FIGURE 18: TAGGING SYSTEM WORKFLOW	53
FIGURE 19: THE WORKING PROCEDURE OF THE PARSER.....	62
FIGURE 20: RESULT OF THE ACCURACY TEST FOR THE CHAT ENVIRONMENT.....	67
FIGURE 21: RESULT OF THE ACCURACY TEST FOR THE ARTICLE ANALYSIS ENVIRONMENT	67
FIGURE 22: PARTICIPANTS' RESULTS FOR STORY ONE	70
FIGURE 23: PARTICIPANTS' RESULTS FOR STORY TWO	71
FIGURE 24: PARTICIPANTS' RESULTS FOR STORY THREE	72
FIGURE 25: THE WORKFLOW OF THE FUZZY EMOTION ANALYSIS FUNCTIONS.....	75
FIGURE 26: AN EXAMPLE OF MOOD SELECTION INPUT	77
FIGURE 27: THE DECAY OF EMOTION OVER TIME.....	78
FIGURE 28: THE COG EXAMPLE	84
FIGURE 29: EMOTIONS EXTRACTED FROM THE EXAMPLE SENTENCE.....	85
FIGURE 30: THE STRUCTURE OF STORAGE COMPONENT	85
FIGURE 31: EXPRESSIVE IMAGE GENERATOR FRAMEWORK	96
FIGURE 32: AN EXAMPLE OF THE SEGMENT AREAS.....	98
FIGURE 33: EXAMPLE OF CONTROL POINTS AND CONTROL AREAS.....	98

FIGURE 34: LOCAL AREA WARPING EXAMPLE	103
FIGURE 35: STRUCTURE OF THE MORPHING SUBSYSTEM FOR HUMAN IMAGES.....	105
FIGURE 36: STRUCTURE OF THE MORPHING SUBSYSTEM FOR CARTOON IMAGES	105
FIGURE 37: GENERATED HUMAN EXPRESSION IMAGES	106
FIGURE 38: THE CARTOON EXPRESSIVE IMAGE SET.....	107
FIGURE 39: HUMAN EXPRESSION IMAGE CLIPS	108
FIGURE 40: CARTOON EXPRESSION IMAGE CLIPS	109
FIGURE 41: OBJECTIVE-DOMINATED ARTICLE	117
FIGURE 42: OPINION-DOMINATED ARTICLE.....	117
FIGURE 43: THE INTERFACE OF THE ANALYSER.....	121
FIGURE 44: THE FTSE-ALL INDEX PERCENTAGE RESULTS	122
FIGURE 45: THE FTSE100 PERCENTAGE RESULTS.....	123
FIGURE 46: FTSE-ALL INDEX CORRELATION BETWEEN DIFFER AND NEXT DAY'S VALUE.....	124
FIGURE 47: FTSE 100 CORRELATION BETWEEN DIFFER AND NEXT DAY'S VALUE.....	124
FIGURE 48: FTSE-ALL INDEX MOVEMENT AND EMOTION MOVEMENT	125
FIGURE 49: FTSE 100 MOVEMENT AND EMOTION MOVEMENT.....	125
FIGURE 50: FTSE-ALL INDEX RESULT OF SMALL DIFFERENCE AND THE NEXT DAY'S MARKET MOVEMENT	126
FIGURE 51: THE EXPRESSIVE CHAT 2D INTERFACE.....	128
FIGURE 52: THE EXPRESSIVE CHAT 3D MAZE INTERFACE	129
FIGURE 53: THE TWO COGNITIVE STYLE DIMENSIONS.....	132
FIGURE 54: VERBALISER AND IMAGER MODE OF REPRESENTING INFORMATION.....	133
FIGURE 55: COGNITIVE STYLE QUADRANTS	134
FIGURE 56: THE MIXED FACTORIAL FULLY CROSSED 2x4 DESIGN	135
FIGURE 57: THE FOUR MEDIA COMBINATION INTERFACES	137
FIGURE 58: THE QUESTIONNAIRE SESSION	138
FIGURE 59: THE CSA SESSION	138
FIGURE 60: HELPFULNESS OF TEXT IN THE E+V+T, E+T AND T+V INTERFACES	143
FIGURE 61: PREFERENCES OF THE COMPONENTS	144
FIGURE 62: SCATTER-GRAPH OF CS	145
FIGURE 63: THE FIELD OF VIEW AND THE POSITION GUIDE IN THE 3D INTERFACE	153
FIGURE 64: DIFFERENT EYE CONTACT ANGLE	154
FIGURE 65: BUSY 2D AND 3D CHAT INTERFACES.....	155
FIGURE 66: THE PERCENTAGE OF INTERFACE PREFERENCE OF THE BUSINESS-ORIENTED ACTIVITIES	159
FIGURE 67: THE PERCENTAGE OF INTERFACE PREFERENCE OF THE SOCIAL-ORIENTED ACTIVITIES	159
FIGURE 68: STRUCTURE OVERVIEW OF AN EMOTIONAL AGENT	167

FIGURE 69: A SCREEN SHOT OF INTERFACE 1 (ONE AGENT)	169
FIGURE 70: A SCREEN SHOT OF INTERFACE 2 (TWO AGENTS WITH NEUTRAL EXPRESSIONS).....	170
FIGURE 71: A SCREEN SHOT OF INTERFACE 3 (TWO AGENTS WITH COMPLIMENTARY EXPRESSIONS)	170
FIGURE 72: A SCREEN SHOT OF INTERFACE 4 (TWO AGENTS WITH OPPOSITE EXPRESSIONS).....	170
FIGURE 73: PARTICIPANTS MEAN RATINGS OF QUESTION 1	174
FIGURE 74: PARTICIPANTS MEAN RATINGS OF QUESTION 2	176
FIGURE 75: PARTICIPANTS MEAN RATINGS OF QUESTION 3	177
FIGURE 76: PARTICIPANTS MEAN RATINGS OF QUESTION 4	179
FIGURE 77: THE WEAR DOWN FACTOR.....	183
FIGURE 78: TYPICAL SCREENS OF THREE CONDITIONS.....	184
FIGURE 79: TYPICAL SCREENS OF THE TWO GROUPS.....	185

List of tables

TABLE 1: THE TAG SUBSET	51
TABLE 2: EXAMPLE OF RECORDED INFORMATION OF EMOTIONAL WORDS	55
TABLE 3: RESULTS OF THE SPEED EXPERIMENT	69
TABLE 4: THE RAW INTENSITY VALUES	79
TABLE 5: THE WEIGHTED INTENSITIES	79
TABLE 6: THE VALUES OF ARRAY E.....	80
TABLE 7: THE "PERCENT" VALUES OF ARRAY E IN THE EXAMPLE.....	80
TABLE 8: OUTPUT OF RULE 1 TO 3.....	82
TABLE 9: THE RAW INTENSITY VALUES	86
TABLE 10: THE WEIGHTED INTENSITIES.....	87
TABLE 11: THE VALUES OF ARRAY FE IN THIS EXAMPLE	87
TABLE 12: THE "PERCENT" MEMBERSHIP'S VALUES OF <i>POS</i> AND <i>NEG</i> IN THE EXAMPLE.....	88
TABLE 13: MEAN SHAPE DESCRIPTON FOR EACH CATEGORY.....	97
TABLE 14: FINISH POINTS VALUE FOR HUMAN EMOTION "HAPPINESS" AND "SADNESS"	99
TABLE 15: FINISH POINTS VALUE FOR HUMAN EMOTION "SURPRISE" AND "ANGER"	100
TABLE 16: FINISH POINTS VALUE FOR HUMAN EMOTION "DISGUST" AND "FEAR"	100
TABLE 17: FINISH POINTS VALUE FOR CARTOON EMOTION "HAPPINESS"AND "SADNESS"	101
TABLE 18: FINISH POINTS VALUE FOR CARTOON EMOTION "SURPRISE"AND "ANGER"	101
TABLE 19: FINISH POINTS VALUE FOR CARTOON EMOTION "DISGUST" AND "FEAR"	102
TABLE 20: THE PERCENTAGE OF THE PARTICIPANTS WHO CORRECTLY RECOGNISED HUMAN EXPRESSIONS IN IMAGE ONLY INTERFACE.....	112
TABLE 21: THE PERCENTAGE OF THE PARTICIPANTS WHO CORRECTLY RECOGNISED HUMAN EXPRESSIONS IN IMAGE PLUS TEXT INTERFACE	112
TABLE 22: THE PERCENTAGE OF THE PARTICIPANTS WHO CORRECTLY RECOGNISED CARTOON EXPRESSIONS IN IMAGE ONLY INTERFACE.....	113
TABLE 23: THE PERCENTAGE OF THE PARTICIPANTS WHO CORRECTLY RECOGNISED CARTOON EXPRESSIONS IN IMAGE PLUS TEXT INTERFACE	113
TABLE 24: THE TAG SUBSET FOR THE STOCK ANALYSER.....	120
TABLE 25: TAGGED DICTIONARY EXAMPLE	120
TABLE 26: PREFERENCES OF DIFFERENT INTERFACES	140
TABLE 27: 2-WAY ANOVA RESULTS OF INTERFACE PREFERENCES.....	141
TABLE 28: HELPFULNESS OF EXPRESSIVE IMAGES IN THE E+V+T AND E+T INTERFACES	141
TABLE 29: CHI-SQUARE RESULTS FOR THE HELPFULNESS OF EXPRESSIVE IMAGES IN THE E+V+T INTERFACE.....	142

TABLE 30: CHI-SQUARE RESULTS FOR THE HELPFULNESS OF EXPRESSIVE IMAGES IN THE E+T INTERFACE	142
TABLE 31: HELPFULNESS OF VOICE IN THE E+V+T AND V+T INTERFACES	142
TABLE 32: CHI-SQUARE RESULTS FOR THE HELPFULNESS OF VOICE IN THE E+V+T INTERFACE	143
TABLE 33: CHI-SQUARE RESULTS FOR THE HELPFULNESS OF VOICE IN THE V+T INTERFACE.....	143
TABLE 34: HELPFULNESS OF TEXT 2-WAY ANOVA RESULTS	144
TABLE 35: IMAGE PREFERENCE BY CS.....	145
TABLE 36: CHI-SQUARE RESULT FOR COMPONENT PREFERENCE.....	146
TABLE 37: THE DIFFERENT CATEGORIES OF ACTIVITIES.....	150
TABLE 38: THE STYLE OF ACTIVITIES	157
TABLE 39: THE SOCIABILITY ASSESSMENT QUESTIONS.....	158
TABLE 40: DEPENDENT VARIABLES.....	160
TABLE 41: QUESTIONS AND THE CHOICES FOR THE ONLINE QUIZ.....	171
TABLE 42: FOUR CATEGORIES OF THE ONLINE QUIZ INTERFACES	171
TABLE 43: THE QUESTIONS IN QUESTIONNAIRE SESSION.....	172
TABLE 44: MEAN RATINGS AND STANDARD DEVIATIONS FOR QUESTION 1	173
TABLE 45: MEAN RATINGS AND STANDARD DEVIATIONS FOR QUESTION 2	175
TABLE 46: MEAN RATINGS AND STANDARD DEVIATIONS FOR QUESTION 3	177
TABLE 47: MEAN RATINGS AND STANDARD DEVIATIONS FOR QUESTION 4	179
TABLE 48: MEAN RATINGS AND STANDARD DEVIATIONS FOR QUESTION 5	181
TABLE 49: QUESTIONS ASKED IN THE QUESTIONNAIRE SESSION.....	185
TABLE 50: DESCRIPTION OF THE EXPERIMENT DESIGN TO TEST HYPOTHESIS 3.....	186
TABLE 51: RESULTS OF THE TESTS OF HYPOTHESES.....	188

Preface

This work was sponsored by Bournemouth University, UK

Acknowledgements

I would like to thank Professor Anthony Boucouvalas for his constant enthusiasm, belief and guidance, Dr. David John for the technical assistance, and Hasan Chowdhury for his important contributions in the early stages of the project.

I also thank my parents, my wife XiaoYuan Xu, and her parents for their long support, encouragement and understanding.

For further details of the software described in this thesis contact:

Research Administrator
School of Design, Computing and Engineering,
Bournemouth University
Fern Barrow
Poole
Dorset
BH12 5BB
UK

Chapter 1

Introduction

This research project - Real-time Expressive Internet Communications focuses on the investigation of automatic emotion detection and visualisation in Internet communication, e.g., emotional feelings expressed in Internet chat rooms or in Internet human-computer interactions. In this context, text is an important communication channel to convey messages and present emotion feelings, thus textual messages containing significant amounts of emotional feelings are transmitted over the network by users.

The Internet has seen tremendous growth, not only with the ranks of new users expanding at ever-increasing rates, but also with the technology revolution. This expansion has catapulted it from the realm of academic research towards mainstream acceptance and increased social relevance for the everyday individual. Users of the Internet around the world browse millions of web pages every day and send out even more textual messages to comment public events, to ask for help or just chat with people they have never met before. Text is an extremely important tool to transmit messages and convey emotional feelings over the Internet. However the feelings embedded in textual Internet messages can not be easily visualised automatically by Internet interfaces. In consequence, various tools (e.g. emotion icons) have been provided for users to manually present their emotional feelings. Most current Internet interfaces do not provide the automatic emotion detection and visualisation feature.

The plan of the project is to investigate approaches of enhancing the sensing of emotion from user's textual messages by displaying corresponding expressive images within the Internet communication environment. A theory to illustrate the emotion communication process over the Internet and an emotion extraction engine that can detect and extract emotional feelings contained in textual messages are also required.

Approaches of synthesising expressive facial images for a user in a user-friendly interface need to be devised. Although numerous different facial expression generation algorithms already exist, most of the algorithms are not suitable for the Internet communication purpose. One reason is the computation time and the other is the complexity of the interface interaction. An expressive image generator with an easy-to-use interface and low computation time consumption needs to be created to synthesise expressive images with acceptable qualities.

1.1 Motivation

Automatic emotion detection and visualisation over the Internet had been steeped in controversy. Within the Internet communication study, researchers were split into two camps, those that believed emotion was an important factor in Internet communication and those that believed that its worth was questionable. However current researchers agree on the importance of the emotion detection and visualisation over the Internet. In this research work, it is assumed that emotion detection and visualisation is a worthwhile research pursuit for three reasons. First, users desire it. In an online communication environment where users can express feelings freely, the needs to express emotional feelings visually to other users are the utmost importance. Second, the methodologies to automatically detect and visualise emotional feelings over the Internet communication have been relatively ignored. Most current emotion presentation methods are based on manual input and the visualised emotion can be as simple as an emotion icon. Third, previous researches have not explored the consequences of detecting emotion and presenting visualised emotional feelings to Internet users.

1.2 Statement of the problem

The first problem this research work attempts to solve is how to automatically detect and visually project emotion from Internet textual messages in order to provide a feature of real-time expressive communication for the Internet. The second problem is to assess the effect of presenting appropriate expressive images to Internet users. In order to explore these issues, the following questions have to be proposed and answered in this project.

- **What is emotion?**

This question provides the fundamental background for other questions. This question can be divided into sub-questions such as the definition of emotion, the categories of emotion and the process of emotion communication etc. To answer this question, literature surveys are carried out and different theories related to emotions are compared.

- **Is emotion an important factor within the Internet communication context?**

This question justifies the meaning of this project. If emotion is not important or even does not exist within the Internet communication environment, the project will be meaningless. To answer this question, not only literature about emotion within computer and Internet

communications are examined, but also experiments assessing the importance of emotion over the Internet are to be carried out.

- **How can emotion be detected and extracted?**

This question focuses on the technical issues. In order to answer this question, firstly literature on Internet emotion will be reviewed and current emotion detection approaches are to be examined. Secondly, an emotion extraction system – emotion extraction engine for Internet communication will be implemented.

- **How to visualise the extracted emotion?**

This question focuses on the technical level as well. Similar to the previous answer, literature on expressive image synthesis are assessed. Based on the literature review findings, an expressive image generator is to be implemented.

- **What are the influences of presenting an interface with the automatic emotion detection and visualisation feature on Internet users?**

This question focuses on the performance assessments and influence assessments of the emotion extraction engine. To answer this question, first literature studies of the possible impacts of emotion will be carried out. With the theory findings, a series of experiments need to be carried out to assess the influences of automatic emotion detection and visualisation.

1.3 Methodology of this study

This section concentrates on the methodologies of this study. How can individual's emotion be detected and visualised in real-time Internet communication? How to assess the influences of automatic emotion detection and visualisation over Internet communication? At a very general level, literature reviews, prototype developments and experiments are the answers to the above questions.

- **The literature review methodology**

First of all, general literature searches need to be carried out to gain knowledge of the current status of the problem this project aims to solve. Internet search engines (e.g. Google, Yahoo and

MSN) and individual periodical databases (e.g. SiteSeer and InfoTrac) will be used as the main search tools and Internet and peer reviewed journals will be the major sources to gain knowledge. Published books will be mainly used to gain fundamental knowledge on theories, mathematics and software programming languages.

- **The prototype development methodology**

The requirements of the prototype need to be discussed with internal users (e.g. colleagues) and external users (e.g. students). The limitations of the Internet environment should be considered as well. Then software development languages will be chosen to meet the requirements. The windows platform is selected to be the main running platform for the prototype although the prototype should be able to be deployed into other operation systems.

- **The experiment methodology**

Various experiments are required to be carried out to assess the performance of the prototypes and examine other related issues. As this research aims to provide emotion visualisation features for Internet communication, some experiments may be carried out through the Internet. Questionnaires will be used to collect feedback from users. Human factors (e.g. human cognitive style and sociability) will be treated as dependant variables when assessing the prototypes in order to analyse the prototype more precisely. Statistic analysis will be widely used in the experiments to determine the significance of the experiment results. Various statistic techniques (e.g. t-tests, ANOVA tests and correlation tests etc.) are used to analyse the experiment data and SPSS (statistical analysis software developed by SPSS Inc.) is chosen to be the main statistical analysis software. The significance value of less than 0.05 ($p < 0.05$) indicates that a tested hypothesis is correct. When the significance value is between 0.05 and 0.10, the result indicates that there is marginal evidence that the hypothesis is correct and the significance is called marginal significance [Schuyler 2003].

1.4 Outline of this thesis

The thesis is structured in the way that reflects the different stages of this research project. This chapter, chapter one, describes the research problems, the methodologies to solve the problems and contributions to knowledge of the research work.

Chapter 2 reviews literature relevant to the background of the problems to be solved. A number of emotion theories and emotion detection algorithms are discussed, and how they can be applied to different contexts is briefly reviewed. It is argued that the emotional feelings can be detected in text, speech and facial movements. This research work is particularly interested in the investigation of emotional feelings embedded in text messages. As a result, different emotion sensing approaches for textual messages are presented and numerous theories and algorithms of synthesising facial images are also introduced.

Chapter 3 first presents the emotion theory that was followed and discusses the emotion momentum theory that was developed to illustrate the emotion communication process over Internet communication. Then an overview of the prototype system architecture is presented and the two components of the prototype system - emotion extraction engine and expressive image generator will be described in brief.

Chapter 4 discusses the emotion extraction engine in detail. First, the development of the emotion extraction engine, including the structure of the engine, the components of the engine and rules applied to guide the engine's analysis, is discussed. The second part of chapter 4 presents two experiments assessing the accuracy of emotions detected by the emotion extraction engine. The first experiment inputs sentences from both published articles and from messages written in chat environments into the emotion extraction engine. The emotions detected by the emotion extraction engine are compared with emotions identified manually. The second experiment assesses the emotion extraction engine's ability to handle conflicting emotions.

Chapter 5 first describes the fuzzy logic components of the emotion extraction engine in detail, and then an experiment designed to assess the performance of the fuzzy logic components is presented. In the experiment, three stories containing conflicting emotions are presented to experiment participants and to the emotion extraction engine. The emotions identified by the participants and the emotions sensed by the emotion extraction engine are compared to examine the performance of the fuzzy logic components.

Chapter 6 first presents the architecture of an expressive image generator. The expressive image generator can synthesise not only human facial expressions but also cartoon expressive images. The image generator applies image warping and image morphing techniques to generate expressive images belonging to six expression categories and three different intensities. The second part of chapter 6 presents experiments assessing the performance of the expressive image generator. The experiments presented the expressive images generated by the image

generator to Internet users, and examined the feedback from the users to assess the recognition rate of the synthesised images.

Chapter 7 presents different applications embedding the emotion extraction engine. A 2D-chat environment, a 3D-chat environment, an online-game and a stock market emotion analyser are discussed. These applications show the possibilities of tailoring the emotion extraction engine into different environments. The discussion focuses on the stock market emotion analyser, which is a typical example of tailoring. In the stock market emotion analyser, the tag sets were re-designed and the tagged dictionary was re-created. An experiment assessing the correlation between the emotions detected by the stock market emotion analyser and the movement of the next day's share prices is presented.

Chapter 8 presents an experiment assessing the preference of interfaces embedding the emotion extraction engine. Four different interfaces containing different components (text, voice and expressive image) and different component combinations are viewed by experiment participants. The feedback from participants were analysed to find the preference difference between different component and different combinations. The experiment results demonstrate that users prefer an interface with automatic expressive image display.

Chapter 9 describes an experiment that assessed important social factors which may influence the expressive interface preferences. The assessed factors include interface style, activity style and sociability. The different social feelings perceived in a 2D interface embedding the emotion extraction engine and a 3D interface embedding the emotion extraction engine are compared. As the 2D and 3D expressive interface may be suitable for different activities and individual's sociability style may influence the choice, an experiment was carried out to test the relationships between activity style, interface style and sociability. The experiment shows that people prefer to carry out more casual and relaxing activities in the 3D expressive interface and choose the 2D expressive interface to carry out less sociable activities.

Chapter 10 first presents an experiment that detected the influence of presenting expressive images to users. The experiment results illustrate that expressive images do significantly influence the feelings of users and affect their preferences of interface styles. Then an experiment assessing the different factors that may influence the presentation of expressive images is presented. The assessed factors include expression intensity and timing effort. The results demonstrate that expression intensity does significantly influence the perceived

emotional feelings and performance; however the influence of timing effort is not supported by the experiment.

Chapter 11 presents the conclusions of this research project and further research suggestions. Recommendations on how to detect emotions contained in textual messages and how to improve the performance of the user interface by introducing expressive images are discussed. The internal and external factors that may influence the perceived emotional feelings are reviewed. The suggestions for further research of similar areas are given.

1.5 Contributions to the knowledge of the research work

The contributions to knowledge of this research work fall into four main areas: First, the proposed emotion momentum theory for the Internet communication environment. Second, the innovations built into the prototype emotion extraction engine. Third, the approach developed to generate facial expressions for individual Internet users in a user friendly and fast manner. Fourth, the identification of the influences of visualised emotions on the perception of emotional feelings within human-computer interaction.

The emotion momentum theory is proposed to illustrate the emotion process within the Internet communication context. The contributions of this theory include:

- The identification of the three internal factors of the communication of emotion over the Internet: emotion category, emotion intensity and duration.
- The identification of the relative influence of each communicated emotion to other sentences.

The prototype emotion extraction engine and expressive image generator are built to automatically extract and visualise emotions. Existing usability principles, which are discussed in chapter 2 and methods used in Internet communication systems are applied in the emotion extraction engine and the expressive image generator. However, the emotion extraction engine and the expressive image generator are unique and contain a number of innovations associated with the technology used to build the system, how information is structured, and the design of the style of interaction:

- The design of the screen layout. The design of the objects on the screen and the particular mixture of media used to present information and methods of receiving input from the user.

- The user friendly interface of the expressive image generator.
- The use of fuzzy logic to calculate emotional feelings in the emotion extraction engine.

Another major contribution to knowledge is the identification of the influence of emotion visualisation on Internet users:

- The identification of user preference for the emotion visualisation interface: experiment results demonstrate that users significantly prefer interfaces with the visualised presentation of emotion more than without.
- The correlation found between the emotions expressed in stock market articles and the next days share price movement.
- The identification of the influence that the visualisation of emotion has on the feelings perceived by users: users treat computers as humans, thus social norms in daily life are applied in human-machine interaction. When computers present emotion to a user, the emotion will either negatively or positively influence a user's perception of own performance.
- The identification of the internal factors, such as expression intensity and display duration, may influence the perception of emotional feelings.

1.6 Limitations of the study

The literature searches are conducted through periodicals dated between 1980 and 2004 except emotion theories that were developed before that period, and only through those accessible electronic periodical databases, i.e., InfoTrac, MAS (Magazine Article Summaries Full Text Elite) and SiteSecr. This eliminates the historical perspective and the need for information published in literature prior to that time. Most of the emotion literature gathered will be pertinent to Europe and North American, and may not be transferable to other areas of the world, especially those with different emotion cultures and religions, etc.

The emphasis of this study is on the Internet chat communication environment rather than other Internet applications. With the huge amount of research papers and information being published daily through conventional methods and the Internet, some of the information used for this study may be outdated by the time of publication.

1.7 Chapter conclusion

This chapter laid the foundations for the thesis by introducing the motivation of the research work. Then the problems of this research aimed to solve were discussed, the methodology was briefly described and the limitations are declared. On these foundations, the thesis can proceed with a detailed description of the research. In next chapter, the results of literature review will be presented. Different emotion theories, emotion detection techniques and emotion visualisation approaches will be examined. The advantages and disadvantages of different theories will be discussed as well.

Chapter 2

Background review

This project focuses on automatic emotion detection and visualisation in real-time Internet communications. The area of the research covers emotion theory, emotion communication over the Internet, emotion detection, emotion visualisation, human factors and other issues (e.g. HCI principles etc.) relating to the assessment of the emotions. In this chapter, the results of the background study are presented.

2.1 Emotion theories

Individuals present emotional messages and show emotional behaviour to express their feelings. By observing an individual's expression and behaviour, the internal emotional feelings may be revealed in some aspect. When someone is shouting, we may describe him or her as 'angry'. Emotion is a frequently used word; however the detailed mechanisms are not yet revealed. We do not know what kinds of condition generate what emotions and what behaviour trigger which particular emotions. Neither psychological knowledge of emotion nor medical research can explain the mechanisms underlying emotion. The only current method that tells about the mechanism is brain wave analysis, which conceptualises that right brain is in charge of emotion activities [Evatt 1997].

Despite the fact that the fundamental biological process has not been fully explored, emotion research has been a popular field of study that can be traced back to the time of Plato. Numerous theories have been developed to illustrate the process of emotion and the influences on human behaviour. In this section, the five most popular emotion theories: James-Lange theory, facial-feedback theory, Canon-Bard theory, Schachter-Singer theory and cognitive appraisal theory are briefly discussed.

2.1.1 The James-Lange theory

The James-Lange theory, which was proposed in 1884, combined ideas of William James and Danish physiologist Carl Lange. The theory described by James [1878] is that the perception of external facts directly influences bodily changes and feelings.

The James-Lange theory specifies that physical reactions come first when an event happens [James 1878]. This theory proposes that emotions are triggered as a result of physiological events, rather than being the cause of them. In the James-Lange theory, the participative experiences of emotions are nothing more than the awareness of our bodily changes in the presence of certain arousing stimuli [Cannon 1927]. Figure 1 illustrates the general process of the James-Lange theory.

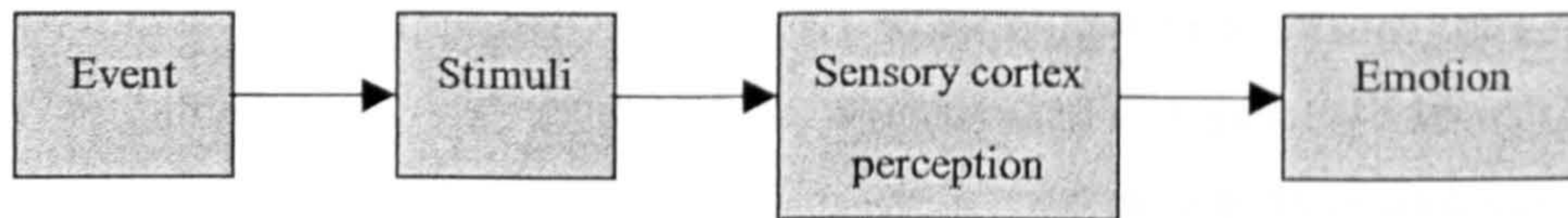


Figure 1: The process of the James-Lange theory

The James-Lange theory is generated from introspection and no experiment results have yet supported it. Most researchers have ruled out this theory as emotion still occurs where the stimuli are small and it is extremely difficult for people to be aware of the physical change [Lange and James 1922].

2.1.2 The Cannon-Bard theory

Cannon and Bard opposed the James-Lange theory by stating that the emotion is felt first, and then actions follow from cognitive appraisal [Gross et al. 2000]. The Cannon-Bard theory claims that information sent to the cortex produces emotions at the same time as the physiological and behavioural responses are produced. However, these are independent of one another [Gross et al. 2000]. The Canon-Bard theory is briefly summarised in figure 2.

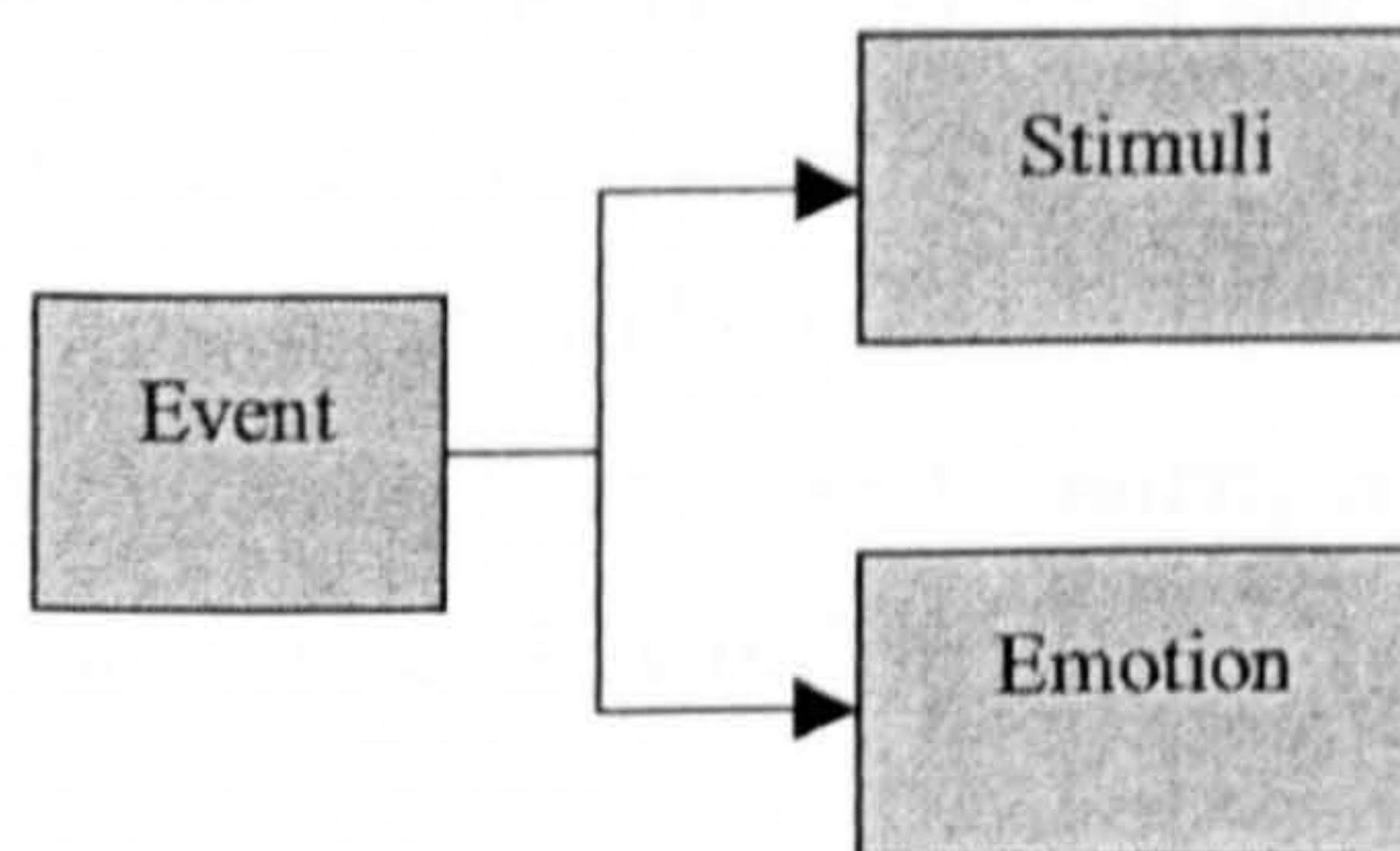


Figure 2: The process of Cannon-Bard theory

2.1.3 The facial-feedback theory

The facial-feedback theory is initially developed by Tompkins [1963] and enriched by Ekman and Friesen [1971] and other researchers. The facial-feedback theory hypothesises that facial expressions can contribute to body feelings and cognitive experiences of the associated emotion.

Certain stimuli are generated in different facial expressions and can produce physiological changes. According to facial-feedback theory, first we smile, then we experience pleasure; first we frown, then we experience sadness. It is the changes in our facial muscles that cue our brains and provide the basis of our emotions. Just as there are an unlimited number of muscle configurations in our face, so too are there a seemingly unlimited number of emotions [Tompkins 1963]. The facial-feedback theory is similar to some aspects of James-Lange theory except that it is the facial muscles instead of the sensory cortex that are responsible of stimuli perceptions. Figure 3 illustrates the general process of the facial-feedback theory.

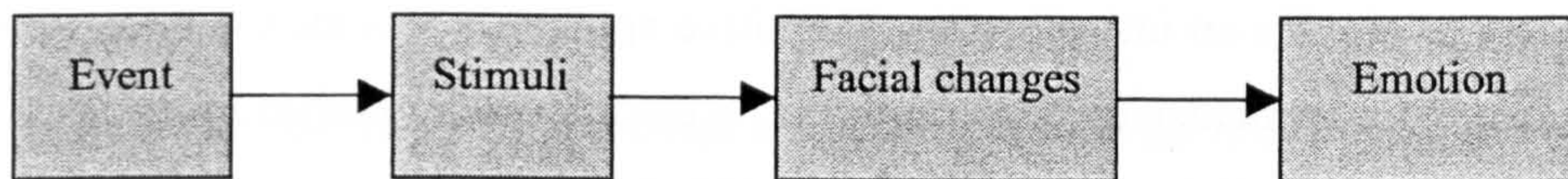


Figure 3: The process of the facial expression theory

The facial-feedback theory is challenged by different studies. Earlier research by Cannon demonstrated that destroying bodily responses does little to affect emotion and emotional displays are not influenced even when the entire sympathetic division of the autonomic system in cats were destroyed [Cannon 1927 and Cannon 1932]. Additionally, studies conducted by Bermond and his colleagues [Bermond et al. 1991] suggested that individuals with spinal cord damage have few bodily correlations of emotion and still reported the presence of emotions.

2.1.4 The Schachter-Singer theory

Schachter and Singer adopted the logic of the James-Lange theory. Schachter [1971] reported that injection of epinephrine did not cause any emotion in particular, but it increased participative reports of the intensity of different emotion categories. The Schachter-Singer theory argued that emotion was determined by cognitive thought, but the intensity of the emotion was determined by bodily responses. According to this theory, an event causes physiological arousal first. You must identify a reason for this arousal and then be able to experience and label the emotion [Schachter and Singer 1962]. The following scenario illustrates this process: You are watching a scary movie late at night. You hear footsteps behind a future victim and see a man with a knife behind. Your heart beats faster, and your breathing deepens. Upon noticing this arousal you realise that comes from the fact that the victim is being attacked. The attack is dangerous and therefore you feel the emotion of fear. Figure 4 shows the general process of the Schachter-Singer theory.

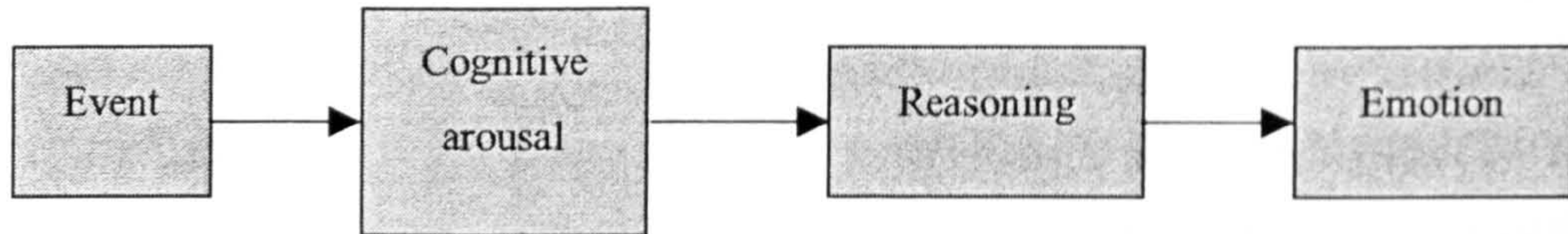


Figure 4: The process of the Schachter-Singer theory

2.1.5 Cognitive appraisal theories

Cognitive appraisal theories represent the leading force in emotion research. In these theories, emotional responses represent undifferentiated physiological states, and cognition is therefore necessary to provide an interpretation. Cognitive appraisals provide the basis for the conscious experience of a particular emotion, and can be used by the organism in an adaptive manner to initiate or alter a particular behaviour [Frijda 1986]. Cognition is necessary to remove ambiguity from indistinct emotional states, and cognitive constructs such as perceptions, thoughts, beliefs, and goals influence this process. The sequence of events involved in an emotional response is described as follows. A stimulus is detected and causes a state of bodily arousal, which in turn is interpreted by the cognitive apparatus to generate an appraisal, which takes into account the individual's goals, plans, and beliefs [Frijda 1986]. Figure 5 illustrates the general process of the cognitive appraisal theory.

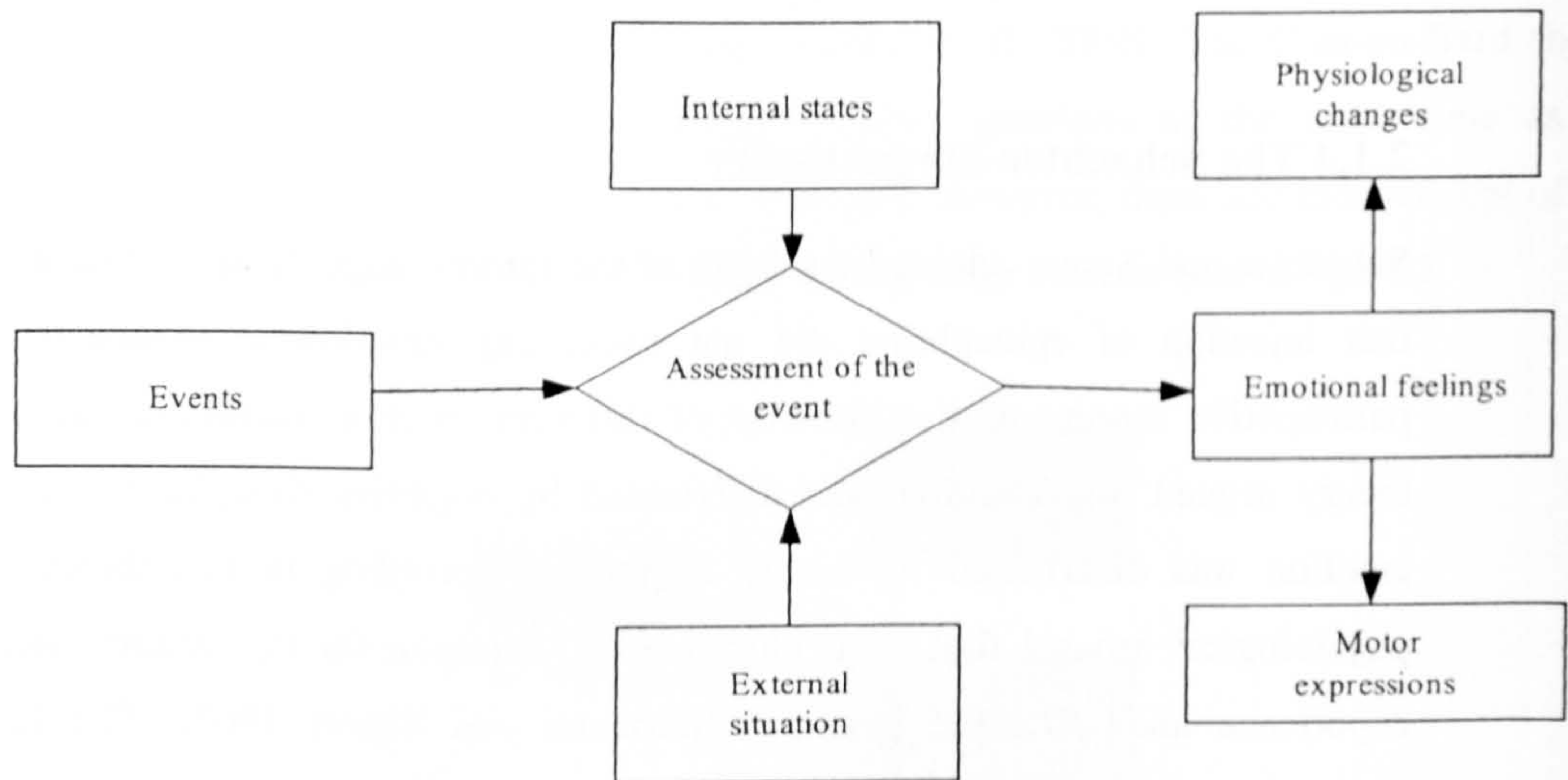


Figure 5: Cognitive appraisal theory

2.2 Classifications of emotion

One of the most well known classifications of Emotion is the Ortony, Clore and Collins (OCC) model, which has established itself as the standard model for emotion synthesis [Ortony et al. 1988]. A large number of studies have employed the OCC model to generate emotions for embodied characters [Andre et al. 1999, Kshirsagar and Magnenat-Thalmann 2002, Adamatti and Bazzan 2002 and Egges et al. 2004]. The OCC model classifies 22 emotion categories and treats emotions as positive or negative reactions to internal or external factors. The reactions are appraised according to the individual's goals, standards and attitudes [Ortony et al. 1988].

The OCC model appears to provide a solid framework for the creation of emotional characters, however it suffered from its own limitations. A summary of problems of OCC model discussed by Bartneck [2002] are presented here: first, the OCC model appears to be too complex for the development of believable characters [Ortony 2001] and it is extremely difficult to map the 22 emotion categories into recognisable expressions. Second, the pure OCC model does not provide a history function which means that characters will totally forget previous emotions. This may trigger an awkward emotion analysis. Assuming a very simple scenario like this: "A cartoon character was hit by you badly and then you give it an apple". Without a history function, the cartoon character would just be happy, but in reality, it should also reflect the anger or sadness of the character. The third problem is that the OCC model is too complex to be suitable for all situations; in fact Ortony acknowledges that the original classification is too complex for the development of believable characters [Ortony 2001]. It is not necessary to use all 22 emotion categories to develop a believable character and only the emotions that can be mapped to facial expressions need be considered [Bartneck 2002 and Ortony 2001].

Ekman proposed six basic expression categories that can be communicated efficiently and across cultures through facial expressions [Ekman et al. 1972, Ekman and Friesen 1978, Ekman and Oster 1979, Ekman 1999 and Ekman et al. 2002]. Ekman's Categories include happiness, sadness, fear and anger which are identified in the OCC model, and also two facial expressions - surprise and disgust that are not classified as emotions in the OCC model. The advantage of Ekman's proposal is the distinctive characteristics between different categories, which makes the facial expressions recognisable and can be easily implemented in software programs. The problem of this classification is that some expressions might not be easily classified into any of the six categories and some people may prefer more detailed categorisation.

As the boundaries of different emotions are ambiguous, emotion conflict is inevitable. Fuzzy logic is theoretically best to handle situations without clear boundaries. In the next section the fundamental elements of fuzzy theory will be discussed.

2.3 Fuzzy logic

Human emotion is an extremely complex process and therefore conflicting emotions widely exist. For example, an individual may feel sad, happy and even fearful simultaneously. Even if someone is in an extreme mood (e.g. extremely happy), he/she may still present other emotions (e.g. surprise) simultaneously. As emotion does not have clear-cut boundaries and individuals' emotions are changeable, fuzzy logic is theoretically best for detecting conflicting emotions.

Dr Lotfi [1965] developed the idea of fuzzy logic. Fuzzy logic is a superset of Boolean logic that has been extended to handle the concept of partial truth - truth values between "completely true" and "completely false". It is the logic underlying modes of reasoning which are approximate rather than exact. Instead of working out the exact value in Boolean logic, fuzzy logic is viewed as a limiting case of approximate reasoning. The fuzzy logic is based on fuzzy subset and the system based on fuzzy logic is called a fuzzy expert system. The suitability for applying fuzzy logic into emotion analysis area derives from the fact that most modes of human reasoning and especially common sense reasoning are approximate in nature. In this section, fuzzy subset and fuzzy expert systems are described in brief.

2.3.1 Fuzzy subsets

A fuzzy subset (F) can be defined as a set of ordered pairs, each with a first element that is an element of the whole set (S), and a second element that is a value in the interval [0, 1]. This defines a mapping between each element of the set S and values in the interval [0, 1]. The value zero represents complete non-truth, the value one represents complete truth, and values in-between represent intermediate degrees of truth. The set S is referred to as the universe of discourse for the fuzzy subset F [Bezdek 1989]. Frequently, the mapping is described as the membership function of F, and the values are described as the degree of membership [Bezdek 1989].

2.3.2 Fuzzy expert system

A fuzzy expert system [Driankov et al. 1993] is an expert system that uses fuzzy logic instead of Boolean logic to carry out data analysis and the decision making process. As a result, a fuzzy

expert system is a collection of membership functions and rules that are used to reason about data. Unlike conventional expert systems, fuzzy expert systems are oriented toward numerical processing. The first operation step of the fuzzy system is to define rules and membership functions. The second step is to apply this knowledge to specific values of the input variables to compute the values of the output variables [Bezdek 1989].

Fuzzy expert system is a combination of four processes: fuzzification, inference, composition, and defuzzification. In the fuzzification sub process, the membership functions defined on the input variables are applied to their actual values in order to determine the degree of truth for each rule premise. In the inference sub process, the truth value for the premise of each rule is computed, and applied to the final decision of each rule. This results in one fuzzy subset to be assigned to each output variable for each rule. In the composition sub process, all of the fuzzy subsets assigned to each output variable are combined together to form a single fuzzy subset for each output variable [Bezdek 1989].

Frequently, the result of the composition process needs to be converted to a single number - a crisp value, which is what the defuzzification sub process does. Mizumoto and Zimmerman [1982] compared roughly thirty defuzzification methods. Two of the most common techniques are the CENTROID and MAXIMUM methods. In the CENTROID method, the crisp value of the output variable is computed by finding the variable value at the centre of gravity of the membership function for the fuzzy value. In contrast, the MAXIMUM method chooses one of the variable values at which the fuzzy subset has its maximum truth value as the crisp value of the output variable [Bezdek 1989].

The fuzzy expert system is an important way of representing knowledge imprecisely. The distinct advantage of the fuzzy system is the representation of concepts in a manner that as it is relatively easy to understand for humans. However, with the increasing complexity of fuzzy expert systems, one disadvantage is that its representation power accuracy decreases (e.g. an aircraft control system) [Qian et al. 2002].

2.4 Emotion communication over the Internet

Emotion is ubiquitous in our daily life and is an important feature of human beings. Does emotion exist in computer communication or so-called virtual life? What methods are available

to detect and visualise emotions and what kind of influence would the display of expressive images has on Internet users?

With the expansion of the Internet, many more people intend to spend their time online. Individuals may read online news, search for new technologies and chat with others. The Internet has emerged from a scientific calculation tool to a much larger social environment, thus communicating through the Internet is referred as computer mediated communication (CMC) [Lea and Spears 1991].

Picard [1997] and Bates [1992] demonstrated the importance of emotion over CMC. Picard [1997] demonstrated the potential and importance of emotion in human-computer interaction. Bates [1992] illustrated the roles that emotion plays in user interactions with synthetic agents. In his paper, OZ, an action system based on emotion model, was built to guide the agents' behaviour.

Traditional media (television and newspapers etc.) share communication channels (text, image and animation etc.) with computer communication. Research into the effects of news articles on the readers has developed a theory called "agenda setting". The agenda setting theory asserts that a constant display of images to public through newspaper and television channels shape the priority and attention people give to certain issues [Evatt 1997].

2.4.1 The importance of emotion communication over Internet interaction

In the nineteen eighties, it was thought that the Internet was unsuitable for establishing relationships as text was the only channel that was available to convey information. These assumptions brought about the belief that CMC is "less friendly, emotional, or personal and more business-like, or task oriented" [Rice and Love 1987].

However, social communication researchers [Short et al. 1976 and Bailenson et al. 2001] argued that even in a text-dominated environment, social feelings not only existed but provided important functions. Nowadays, new communication channels (e.g. images, video, audio and facial expressions) have been introduced into CMC and may strongly increase the social feelings perceived in human-computer communication.

Scott and Nass [2002] demonstrated that humans extrapolate their interpersonal interaction patterns onto computers. Humans talk to computers, are angry with them or even make friends

with them. This is the evidence that social norms applied in our daily life are still valid for human-computer interaction. Furthermore, emotion visualisation in the human-computer interface could significantly influence the perceived performance and feelings of humans.

2.4.2 Emotions in believable agents

Believable agents are a popular field for artificial intelligence research. To achieve the concept "believable", many models have been developed to simulate human minds, especially the human emotion processes. The earliest model can be traced to 1967 [Simon 1967]. Most models are based on motivational states and event triggers. It is reasonable to assume that a software agent's emotion is neutral at the starting stage. As time goes on, some events, e.g. being hit, may influence the emotion. If the hit stings with low intensity and occurs just one time, the agent's emotion will be moved a little towards the emotion anger but stays within the boundary of neutral. If no other events were triggered, the agent's emotion will move back to the centre of the neutral. However when the event hit with low intensity stings continuously or stings with a high intensity, the emotion will move into anger and stay if no other events (e.g. to eat and drink) are triggered.

Although human emotional processes are much more complex than above example and it is difficult to build a complete computational model, various models and applications have been developed and applied in the human-agent interaction environments, such as the OZ project [Bates 1992], the Cathexis model [Velasquez 1997], Elliot's Affective Reasoner [Elliot 1992] and the FLAME project [El-Nasir et al. 2000].

2.4.3 Internet avatars

Avatar is defined as "an incarnation in human form or an embodiment (as of a concept or philosophy) often in a person" by the Merriam-Webster dictionary. In the context of computing, software avatars are often referred to as artificial intelligence agents that computer users are willing to perceive as believable or life-like [Prendinger et al. 2003]. Avatars often have adaptive interfaces that respond to individual users in a unique manner and show emotional responses.

Although the functions undertaken by different software avatars differ significantly, a common feature of most software avatars is the human-like interface (e.g. a face of human or a face of an alien). A popular method to design avatar interfaces is to apply cartoon images. As most software avatars interact with users directly, emotion expression is an important factor to

improve the human-computer interaction performance. For example, in some circumstances an avatar may smile when some happy events occur, and may cry if the avatar is hurt. Most current avatars have predefined and limited emotion expressions for a specific face. The lack of automatic emotion expression alteration is a clear defect for software avatars. Researchers such as Bates [1994], Trappl and Petta [1997] and Paiva [2000], have highlighted the importance that displaying emotions has in creating believable characters.

2.5 Computational emotions in other scenarios

Emotions can be revealed not only in day-to-day conversations but also in written text. Low and Repin [2002] examined empirical evidence and carried out a series of experiments. The results of these experiments show that emotion plays an essential and indeed a natural and nurturing part of the way individuals build relationships with public concerns. A large number of articles in newspapers contain emotional feelings and a looping effect was documented by Evatt [1997]. The looping effect can be described as this: Emotional content or so-called emotional articles can influence public perceptions, perceptions are reflected in polls, polls influence policy-makers and then those feelings are published in articles and start another turn of the loop [Evatt 1997].

Another type of looping effect of emotion can be found in stock market articles. Share analysts and stockbrokers apply emotional words (e.g. optimistic or depressed) to describe the status of the market. Articles related to stock market present emotional news to reveal the mood of the stock market and may influence the future behaviour of shareholders. The behaviour of shareholders change the market status then the articles publish the market status and start another turn of the loop. A simple stock and user interaction model is drawn in figure 6.

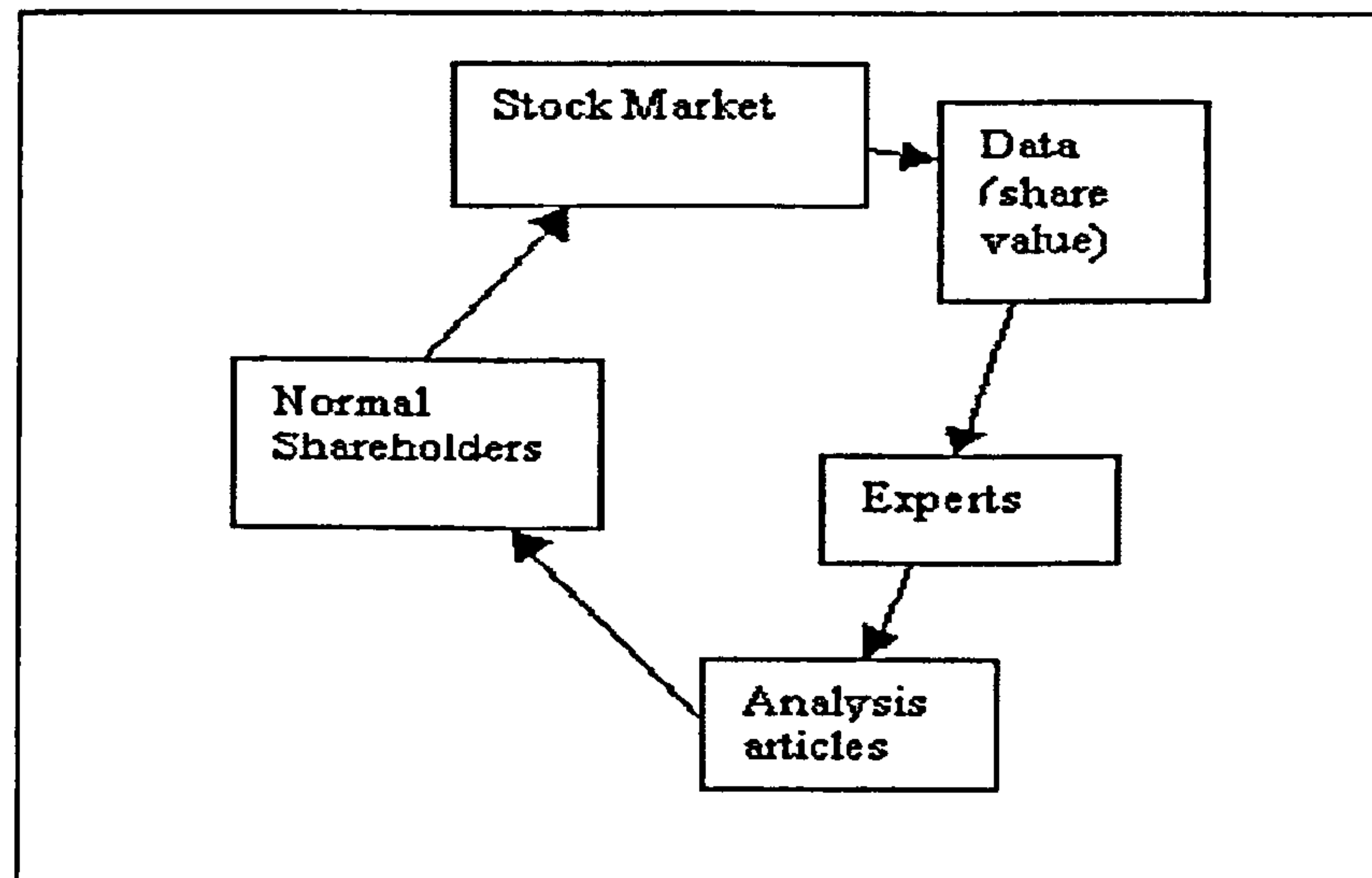


Figure 6: Stock and user interaction

A common stock market phenomenon is that many people buy stocks on a positive news release and sell stocks on a negative news release. The pattern of share price movements and the volume of trading reflect the mood of the market [Rutterford 1993 and Bergen 2003]. Empirical evidence [Evatt 1997, Gaughan 1986 and Nelson 1991] proved the relationship between asset prices and economic news in general. Researchers in MIT found that there is a significant correlation between electro dermal responses and transient market events.

2.6 Emotion detection approaches

It is important to consider the order and process of emotion if one is to attempt to create a valid and reliable measurement methodology for emotion detection. Many researchers have examined this issue and a broad consensus has been reached regarding the order and structure of the methodology. Current methodologies regarding the measurement of emotions can be placed into five main categories, i.e. item based self-report questionnaires, the physiological methods as espoused by Picard and colleagues at MIT [Picard 1997], cross-cultural facial expression detection as espoused by Ekman and colleagues [Ekman 1996], voice expression detection and textual message sensing.

2.6.1 The self-report approach

Self-report is a widely used approach for assessing human feelings. A questionnaire is normally given to each user and users' replies will be mapped into the scales of pre-calculated answers.

Izard [1971 and 1977] and other researchers, e.g. Watson and Clark [1984], took this approach to measure the perceived emotional feelings.

Although widely used by researchers, the self-report method is not suitable for the purpose of this study. First, self-report depends on users' responses and is participative. Second the self-report method requires users to fill in the questionnaire and it is not practicable in the real-time communication context (e.g. a chat room interface). However, the self-report approach is widely used in the experiments of this research work to collect feedback from experiment participants.

2.6.2 The physiological approach

Physiological research is exemplified by the work of Rosalind Picard and her colleagues at MIT. The aims of the work and indeed the original 'Affective Computing' book by Picard [1997] have been the desire to imbue computers or machines with the ability to determine a user's emotion and with an interface to respond in a manner more cognisant which correlates with the users' emotion. In an attempt to achieve this goal, MIT have developed a number of devices, physiological sensors and feature detection algorithms purporting to capture emotionality from physiology [Fernandez and Picard 1997, Riseberg et.al. 1998 and Vyzas and Picard 1999]. The most commonly cited experiment by Picard took sensors to assess each participant's electro-dermal response and blood volume pressure, and found significant correlations between the sensed data and internal affective states. However the result is participant to a host of caveats. The principal weakness is that the participant intentionally elicited the emotions as opposed to the emotion being generated in the normal manner from an internal or external stimulus. Theoretically the physiological method is a good approach as it reveals the internal physiological states. However, practical concerns like cost, usability and the unstable performance forced us to reject this approach in this study.

2.6.3 The facial expression recognition approach

Facial expression recognition systems aim to analyse images of human faces for the purpose of identifying the emotion displayed. A common method of the facial expression recognition is facial templates, which takes a facial image, measures characteristics such as the distance between the eyes, the length of the nose, or the angle of the jaw, and creates a unique file called a "template". The performance of facial expression recognition systems has been reported as successful by different researchers and organizations [Davis and Bobick 1997 and Ezzat and Poggio 1996].

The facial expression recognition approach can be successful in ideal situations (e.g. a lab environment). As the network speed and machines' processing power are still limited for most users, calculating and transmitting high quality facial expressions will be too resource consuming. For this project, instead of recognising facial expressions in real-time, facial expressive images generated automatically in advance will be displayed to represent the emotional feelings.

2.6.4 Voice recognition approaches

A significant amount of work has been carried out in identifying the acoustic features that vary with the speaker's emotion [Arnott and Murray 1993]. By applying different learning algorithms, a number of vocal emotion recognition systems have been developed in the past few years [Dellaert et al. 1996 and Nakatsu et al. 1999]. However, the performance of these recognition systems has not been reliable and some have attempted to use other cues, e.g. facial expression images and texts, to improve their recognition performances [Chen et al. 1998].

This project focuses on the detection and extraction of emotional feelings from textual messages. One reason is the disadvantages of other emotion detection approaches outlined above, another is that text is still the most important method to convey messages and feelings through the Internet. In the next section, different approaches for textual information extraction will be discussed.

2.7 The methods for textual emotion extraction

To extract emotions from Internet communication, various approaches exist, e.g. textual message analysis, expressive image recognition and emotional voice recognition etc. However none of the above methods achieves perfect performance. Textual message analysis was chosen as the main emotion extraction approach based on the following two considerations. First, text is an extremely important method to transmit information and feelings within Internet communications, and is widely used in Internet communications nowadays. Although arguable, text will continue its contribution to the Internet social communication. Second, human language analysis is an extremely complex subject and none of current methods achieve stable performances. However, different text analysis or so-called natural language analysis methods were developed and deployed in some specific contexts (e.g. AI agent character communication etc.). By limiting the context, textual analysis may achieve reasonable performance.

Current approaches of emotion extraction from textual messages can be classified into the following categories: key-word tagging, lexical affinity, statistical natural language processing and real-world knowledge processing. However the fundamental technology that most textual analysis approaches are based on is rewrite-rule analysis.

2.7.1 Rewrite-rule analysis

In human language, sounds and letters make up important parts of the communication. Words are groups of letters and may be combined in many ways to make up a sentence. According to Chomsky [Robinson 1975] and Saussure [Gordon 1996], sentences can be broken down into two groups of words: function words and content words. Content words include nouns, adjectives, verbs and adverbs, while function words are prepositions, conjunctions and auxiliary verbs. All these groups contain sub classes, which can be represented by tree diagrams. In figure 7, an example tree presentation for nouns is unveiled.

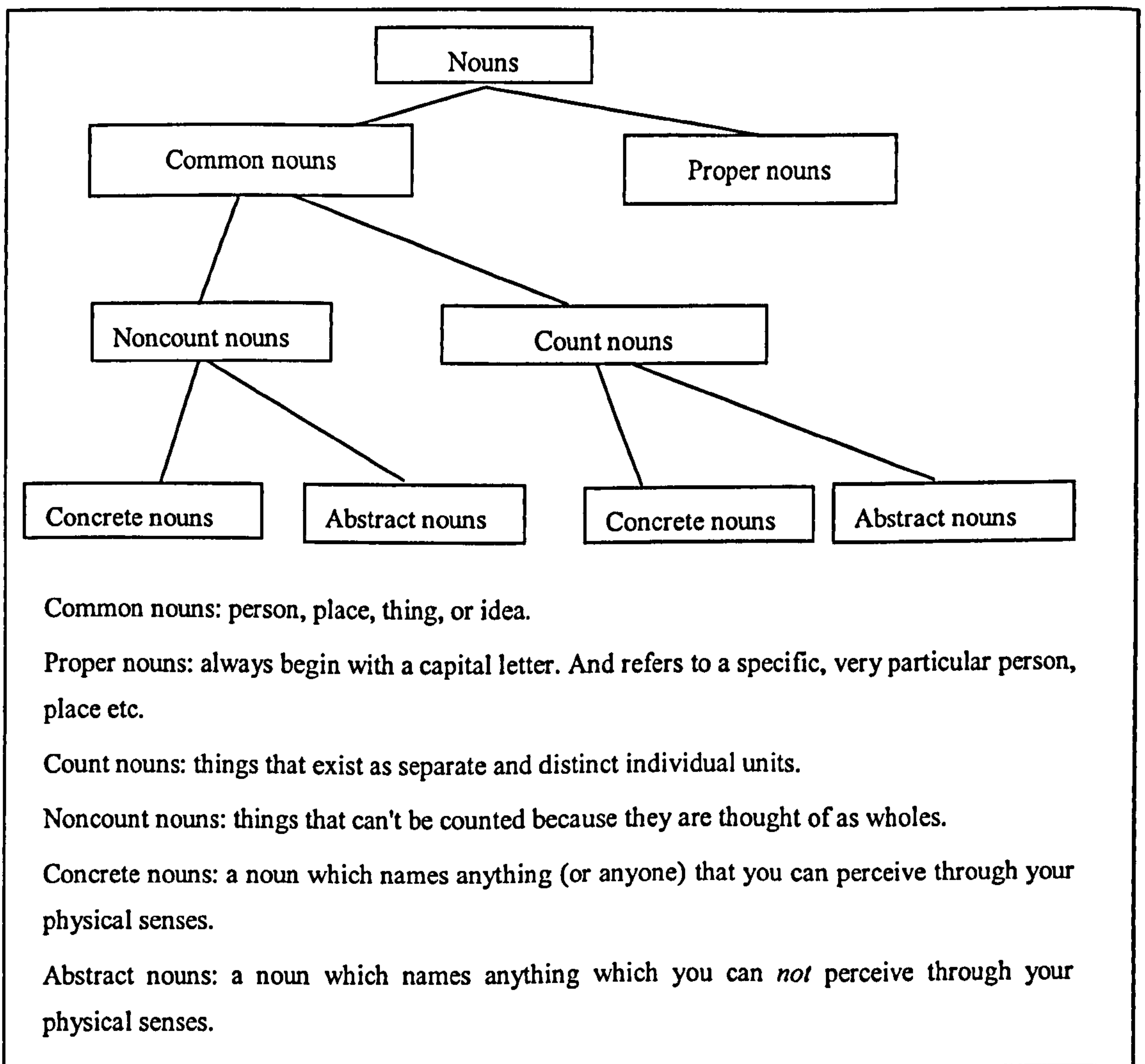


Figure 7: Tree presentation for nouns

A sentence may be decomposed and analysed according to the word categories in the sentence. A sentence may consist of only a single word or a group of words. In order to analyse sentences, a number of generic sentence definition groups have been identified. One example of the sentence definition groups is: "Article + noun + verb + preposition". General sentence constructions have been identified by Crystal [1996]. The constructions include declarative statements, interrogative statements, finite verb phrases and negative statements. The declarative statement usually contains a participant and a verb where the participant precedes the verb, e.g., "I am running". Questions are usually classified as interrogative statements, e.g., "Was he angry?". The finite verb phrases usually express orders or directives, e.g., "Sit down" and the negative statements express the opposite meaning to the content words, e.g., "I am not happy".

Rewrite rules or so-called tree representation [Russell and Novig 2002] are technologies to analyse sentence structures. In figure 8, a tree representation of a sentence is presented.

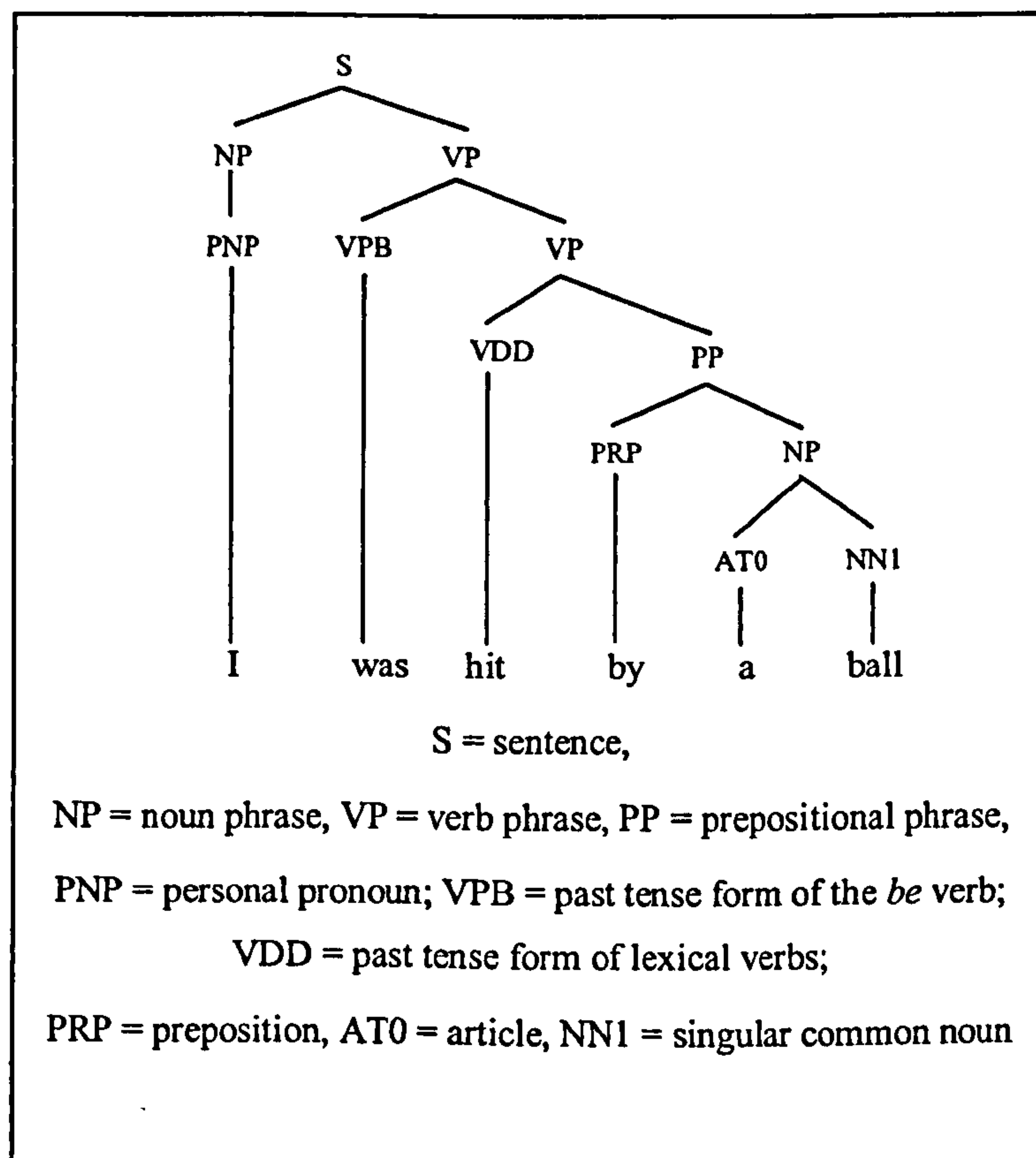


Figure 8: Tree presentation of a sentence

The advantage of rewrite rule analysis is that the analysis process can be easily understood. However, as a sentence's structure can be extremely complex, the rewrite rule analysis may not be able to achieve good performances. Different sentence analysis approaches have been developed based on rewrite rule analysis. In following sections, five of the most important methods are discussed.

2.7.2 Keyword tagging

Keyword tagging is a common approach in textual information analysis. Words are classified into emotion categories based on the presence of emotional feelings. One example of the keyword tagging was the Affective Reasoner [Elliott 1992], which tagged 198 emotional keywords. The emotion keywords can be tagged statically or dynamically according to predefined rules. For example, in sentence "I am happy", the word happy can be tagged as an emotion keyword and an emotion category happiness may be assigned.

The advantages of the tagging approach are numerous, i.e., relatively easy implementation, relatively simple algorithm, fast speed and so on. However, the problem with pure tagging is that some sentences may present emotional feelings without showing an emotional keyword (e.g. "I bought a new car!").

2.7.3 Lexical affinity

Despite the problems of the keyword tagging method, it is still one of the most popular methods of emotion sensing. Based on the keyword tagging method, various extended approaches have been developed. A well known one is "lexical affinity" [Kim et al. 1999]. This approach assigns each emotion word a probabilistic "affinity", e.g. "buy" may be assigned 60% probability of indicating a positive emotion.

This approach presents a flexible way to treat different emotional keywords. However, most implementations of this approach are static based, which means that the probabilistic affinity is assigned previously and the value is fixed. It is difficult to achieve a dynamic probabilistic algorithm and this limits the performance of this method.

2.7.4 Statistical natural language processing

Most statistical natural language processing approaches rely on mathematical descriptions of human languages, and researchers have set up different mathematical models. One model is the hidden Markov model (HMM), which uses different parameters to simulate the language

patterns. To set up a HMM model, keywords, punctuation and common writing phenomena are considered in a statistical manner. The algorithm is trained in a huge corpus (e.g. brown corpus) until it achieves a relative stable performance.

Statistical methods have attracted a large amount of attention; however, statistical methods are generally semantically weak, which means that with the exception of obvious affect keywords, other lexical or concurrence elements in a statistical model have little predictive value individually. As a result, statistical text classifiers only achieve acceptable accuracy by giving sufficiently large text input [Liu et al. 2003].

2.7.5 Real-world knowledge

This approach is based on machine reasoning. Given a large amount of facts, the computer is trained to learn the reasons behind the facts. For example, a fact "bought a car" can be classified as "bought a car → happy" and therefore all sentences with an action of buying a car represents a happy feeling.

Unlike the keyword approaches, this approach is not based on emotion keywords. It presents a totally new and exciting approach, however the approach itself also has inevitable disadvantages. First, the knowledge of the real-world is huge and endless. To implement this approach, designers have to maintain a huge and daily growing database. Second, the real-world itself is ambiguous. A fact that is true in one context may be totally false under other circumstances. For example, the previous fact "bought a car means happy" may be wrong when someone is in poverty but has to buy a car for a reluctant reason. If "bought a car → happy" is deducted from the fact, facts such as "bought a rubbish car" and "bought a second hand car" may both be included into emotion category happiness. However these two sentences do not always present happy feelings. To deal with the ambiguity problem, the only method is to classify the facts in detail, e.g. "bought a rubbish car" means anger and "bought a gorgeous car" means happy. However, this detailed classification approach is essentially equal to keyword tagging as the classification is based on keywords such as "rubbish" and "gorgeous".

2.8 Emotion visualisation technology – facial expression synthesis

There are various facial expression synthesis methods, for example, methods based on modification of facial images and methods based on deformation of a three-dimensional (3D)

polygon model. Facial synthesis has been studied for more than 30 years, yet none of the proposed approaches are 100% realistic.

Most facial synthesis approaches use a learning/training normalization step to create a database. Most systems do the bulk of the analysis work when building the database so that the synthesis step is as easy as possible. Given a particular task, the synthesis step finds the appropriate data in the database, warps it to fit the desired scene, and outputs the final pixels.

3D facial polygon models have attracted most attention in the last ten years. The 3D face models are generic and users do not easily accept the synthesised face. To create a realistic face, animation software (e.g. Maya or 3DMAX) are needed to render the model. It is difficult to achieve the high level performance shown in movies in real time communication. To achieve smooth cinematic performance, it may take more than a week for an experienced designer. However, if the context requires limited number of faces with limited expressions, a cinematic render would be desirable and achievable.

For a human-agent communication scenario, emotional agents should provide expressions with different intensities. In a multi-agent environment, ideally each agent should display individual faces. As a result, the cinematic performance should not be considered, as it requires more than the available resources to achieve. In the human-agent environment, a pre-defined set of faces with different intensities drawn by an artist would be a much more efficient idea as long as that user does not need to create new agents and new facial expressions. According to HCI principles, the operation of user interaction should be as easy as possible. A complex design suite, e.g. FaceWorks [1998], which requires around 100 control points, should not be considered for Internet communication purposes.

For the Internet communication context (e.g. chat room and online games etc.), a diversity of expressions are needed. Users may not only require their own facial expressions with different intensities, but also their favourite cartoon images (e.g. Holmer Simpson etc.) to represent them. Users may prefer to upload a facial picture selected by themselves instead of using pre-selected faces. To generate facial expressions for real-time Internet communication, a fast algorithm and user friendly interface is needed. In the following section, two algorithms – image warping and morphing will be discussed.

2.8.1 Image warping

Image warping is the act of distorting a source image into a destination image according to a transformation between source space (u,v) and destination space (x,y) [Wolberg 1990]. The transformation function f describes the destination (x,y) for every location (u,v) in the source. To apply warping to an image, the transformation function f can be applied to each pixel. Some well-known techniques include radial functions and spline lines. The pseudo code of warping is shown in figure 9.

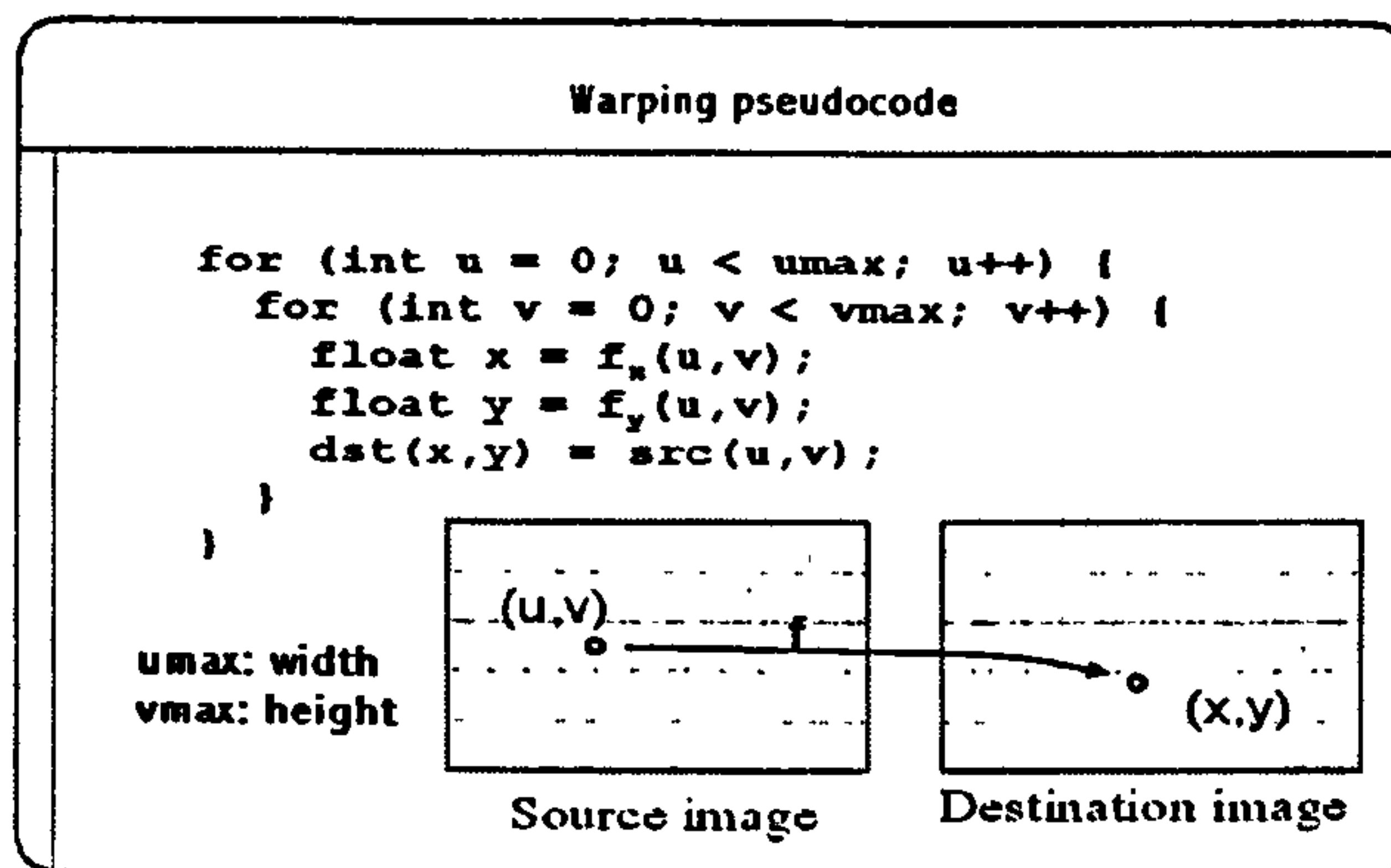


Figure 9: The warping pseudo code

- **Local area warping**

According to Ekman's *Facial action Coding System* [Ekman et al. 2003] and Parke's *Computer facial animation* [Parke and Waters 1996], specific facial muscles are responsible for the generation of expressions. For example, only four muscles contribute to express happiness. As a limited number of muscles on the face are involved in generating different facial expressions, it is much more efficient to carry out manipulation only in those areas.

According to the above analysis, it can be assumed that the noticeable facial distortions are presented in some segmented areas, while in the other part of the image the distortion should be so small that it can be neglected. The segmented area can be chosen as a square that covers the facial distortion. To choose the segmented area, two anchor points (R_1, R_2) are needed. R_1 represents the initial edge and R_2 represents the edge after moving. The width of the segment area is usually chosen as $2 \times \|R_1 - R_2\|$. A typical example of a segment area is presented in figure 10. The point R_1 is chosen to be the right edge of the mouth and R_2 is a point above R_1 , and the segment area is the area covering mouth and nose.

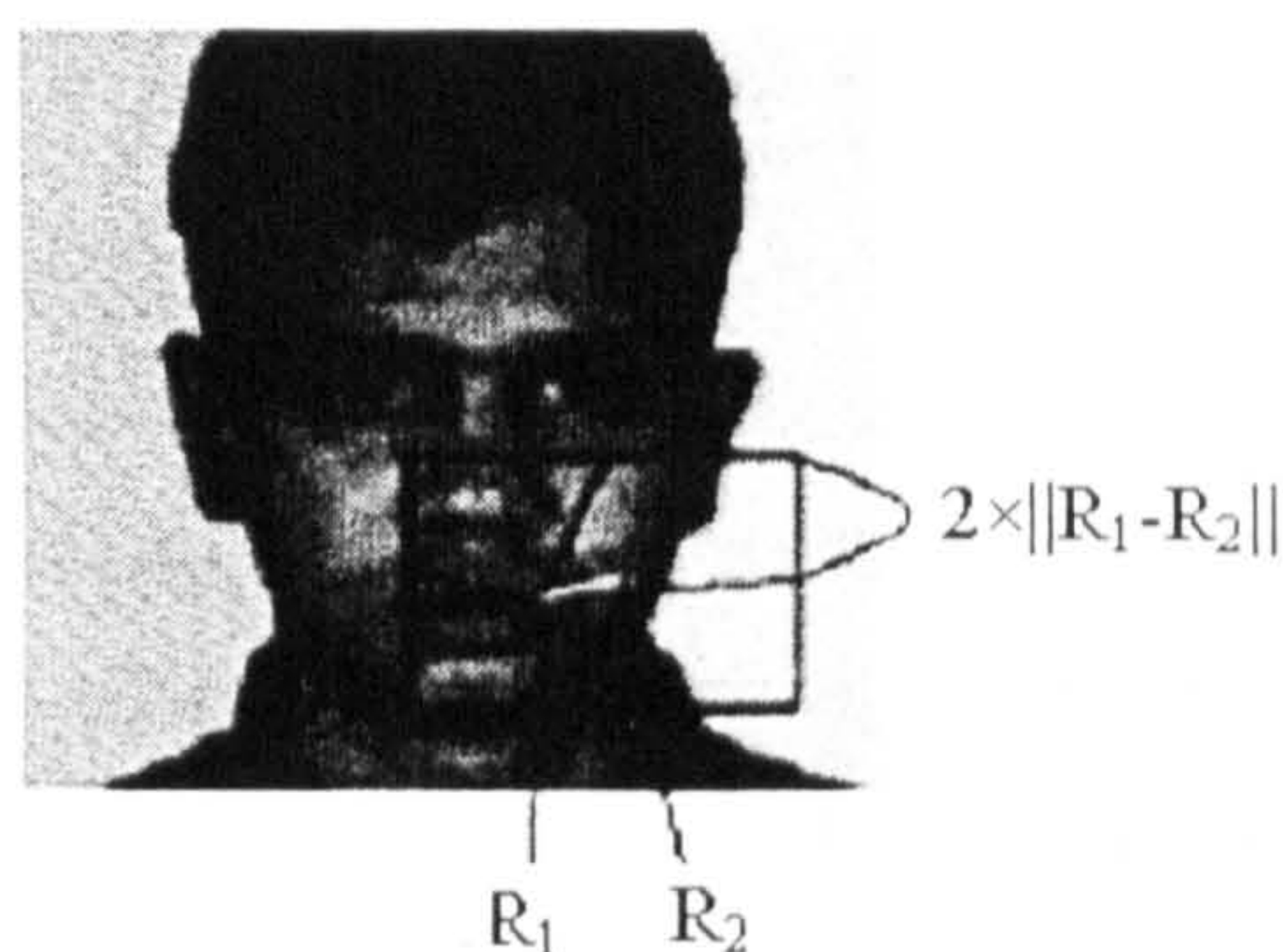


Figure 10: Segment warping area

The advantage of image warping is its fast computation speed [Wolberg 1990]. However, as the warping is calculated on a predefined resolution and each point shows the same amount of influence on the warp [Wolberg 1990], the flexibility of the method are relatively poor. One way to solve the problem is to use local area warping since it sets a limit to the points that can contribute to the warp. Another method is to give a weight to the contributing points. As the weight is generally pre-calculated, the flexibility problem will always exist.

2.8.2 Image morphing

Image morphing is an image processing technique to compute a transformation — metamorphosis, from one image to another. A sequence of intermediate images is created to represent the transition from one image to the other. One of the most famous morphing effects can be observed in the movie “Terminator II”.

In general, there are two different ways of image morphing. Forward mapping scans through the source image pixel by pixel, and copies them to the appropriate place in the destination image. Reverse mapping goes through the destination image pixel by pixel, and samples the correct pixel from the source image. A morph effect can be described as three simultaneous procedures:

1. The shape of the original image, i.e. the "source", is distorted over time to become the same shape as the final image, i.e. the "destination".
2. The final image is distorted to match the "source" image, and it is un-distorted over time until it resumes its normal shape.
3. The opacity of the layers changes over time, so the destination image becomes visible over the source image as they are both distorting.

A significant advantage of image morphing is the computation speed. A disadvantage of image morphing is that the quality of the morphed images may not be very good as gaps may exist between source and destination images. Although methods have been developed to render the gaps and the image, the achieved performance may not always be acceptable.

2.9 Designing easy-to-use interfaces

Numerous emotion extraction and emotion visualisation techniques exist. However, no matter what techniques a system acquires, an easy-to-use user interface is a universal requirement for all systems. Countless books and papers have been published to guide the design of easy-to-use interfaces. A user interface should incorporate features that make the system easy-to-use by implementing the following principles wherever relevant. There are many recommendations of features that can make computer interfaces easy-to-use [Helander 1988, Nielsen 2000, Tognazzini 1990, Mountford and Gaver 1990, Wagner 1990, Dix et al. 1998 and Eberts 1994 and Edwards 1995]. The underlying principles can be summarised in eight golden rules on interface design [Shneiderman 1998]:

1. **Strive for consistency.** Provide a consistent look and feel of the interface, including both the appearance of objects on the screen and the method in which users perform tasks. When the interface looks and behaves consistently it is easier for users to predict the outcomes of certain actions.
2. **Enable frequent users to use shortcuts.** Allow users quickly return to particular places in the system without forcing them through a long route each time. This may involve allowing users to create their own bookmarks to specific places or providing a menu system that enable different routes to the same place.
3. **Offer informative feedback.** Design context-sensitive help information and user instructions that are meaningful and fulfil users' expectations.
4. **Design dialogs to yield closure.** Indicate to users what their current location within the system is and indicate when individual tasks or sections are completed.
5. **Offer error prevention and simple error handling.** Implement an error trapping and reporting mechanism to ensure that the user is made aware if certain operations have failed and do not allow the system to remain in a non-working state.
6. **Permit easy reversal of actions.** Keep a record of the user's route through the system and provide a back button to enable them to retrace their steps and undo certain operations.

7. Support internal locus of control. Enable features that make users feel that they are in control and are not being controlled by the system. Allow users to select the functions of the system they want to use and follow their own route through information.

8. Reduce short-term memory load. Cut complex tasks into short steps or provide adequate instructions so that users can understand the task without having to remember long instructions.

2.10 Chapter conclusion

This chapter reviewed the background information relevant to the development of the theme of this research work. The research work is concerned with creating an easy-to-use method for emotion detection and visualization. Furthermore the research work is interested in investigating the relevant issues such as the preferences of an interface with automatic emotion visualisation and the importance of automatic detecting and visualising emotions. This chapter has explored the nature of emotion, the importance of emotion communication over the Internet and different approaches to the extraction and visualisation of emotional feelings. The next chapter – chapter 3 will focus on the theories relied on as research foundations – the cognitive appraisal theory mentioned in this chapter and emotion momentum theory proposed by this research work. The proposed system architecture to extract and visualise emotion will be explained as well.

Chapter 3

The proposed theory and prototype system overview

The last chapter described different theories about emotion, emotion detection, emotion visualisation and other related aspects. In this chapter, the emotion theories followed, the emotion momentum theory that was developed, and the architecture of the prototype system to sense and visualise emotion is presented.

3.1 The emotion theory this study followed

In chapter 2, various emotion theories are discussed. The advantages and challenges of each theory are presented as well. The agreements found in the literature on emotion theories are summarised below:

- Emotions have valence, which is either positive or negative.
- Emotions are largely triggered spontaneously by some outside stimulus but are then participated to individual processing and modification.
- Emotions are recognised and experienced as such.
- Emotions tend to be volatile and change with external and internal conditions and may, in fact, change in intensity during a single reception.

3.1.1 Cognitive appraisal theory

In this research work, cognitive appraisal theory is followed. The first reason is the popularity – cognitive appraisal theory is the leading force in analysing emotion and other personal feelings. The second reason is the implementation benefit. The cognitive appraisal process can be mapped into the communication scenario in a relatively natural way. The emotion process of the cognitive appraisal theory has been presented in figure 5 in chapter 2, and a summary of the process is discussed in the following paragraph for convenience.

In cognitive appraisal theories, events are raised and assessed by human brain. The assessment – the appraisal is influenced by both internal and external situations. The assessment may lead to the generation of emotion responses and the emotion responses may trigger physiological or physical changes. The emotion process in cognitive appraisal theory can be mapped into the Internet communication scenario as following:

1. **Events.** Theoretically events can be external (e.g. receiving text chat messages from others, movement of the opponent in a game or responses from a software agent) or internal (a thought, memory, fantasy or other effect). The internal events are difficult to detect in the computer communication context and thus the external events are focused on.

2. **Assessment (appraisal).** An assessment of the precipitating event or some degree of cognitive processing can come either just before the experience of the affect, just after the activation of the affect, or exists at both positions. This appraisal can be fleeting or detailed, deeply realistic and empathic. The assessment is influenced by internal states and external situations (e.g. relationship, weather etc.). In the computer and Internet communication context, the internal states can be referred to the internal feelings triggered by the dialog and the interaction process. The external situations are other affecting factors, such as the mood before chatting etc.

3. **Feelings.** Emotional feelings are participative experiences along an intrinsic pain/pleasure axis associated with various perceptions, ideas, sensations, actions, or other dimensions of the precipitating [Morris 1998]. For Internet communication, this may be mapped as the perceptions of the received emotions.

4. **Motor.** Motor, especially facial-motor changes are triggered in emotion process. For a computer communication scenario, this can be illustrated as the expressive image display, expressive voice, or textual emotional messages.

5. **Physiological changes.** Physiological changes include various parameters, skin conductance, and various muscle responses. It is hard to directly detect the physiological changes of a user under the Internet communication circumstance. However, this may be mapped into the emotion transform or current mood transform of an individual user.

3.1.2 Ekman's universal expression categories

Ekman's universal facial expression categories are followed in this study because of the distinctive features between different expression categories. It is certain that Internet communication requires a much richer set of expressions than the six; however the practical consideration – “recognisable expression” is more important than other theoretical considerations. To map emotion into expressions in an easy manner, the emotion categories are chosen to be the same as the six universal expressions, which are happiness, sadness, fear, anger, disgust and surprise.

The cognitive appraisal theory explains the emotion communication process in a general level and Ekman's definition gives the characteristics of six different facial expression categories. To describe the communication of emotion over the Internet, an emotion momentum theory was developed. The theory is based on the observation of Internet chat and study of other research works, e.g., the OCC model.

3.2 Emotion momentum theory

Most modern emotion theories use categories and intensities to represent the emotional feelings, e.g. emotion category happiness and emotion category sadness. The boundaries between different emotion categories are not clear-cut (e.g. we do not know the exact amount of effort needed to push an individual from happiness to sadness). The intensity of a specific emotion category is continuous and the value can vary from zero to extremely high (e.g. from 0 to 3 if 3 is the maximum value of the intensity). For the Internet communication context, the duration of the influences of emotions is uncertain. The emotion momentum theory presents the emotion communication into three dimensions – the emotion category, the duration of an emotion expression and the emotion intensity, which are described visually in figure 11.

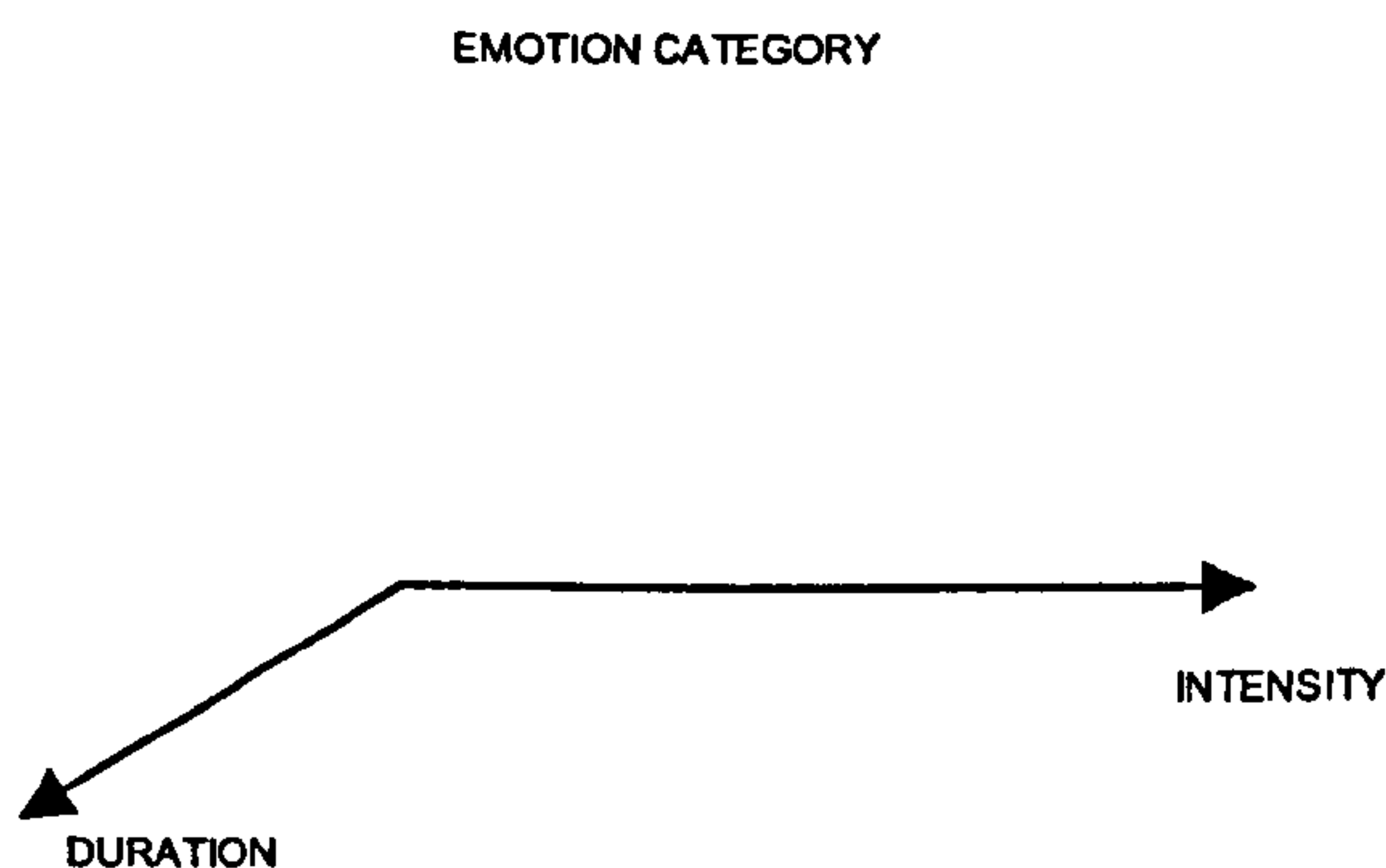


Figure 11: Emotion momentum

The observation is made that at any point in time, people stay in certain emotional states. When interacting with the environments or with other people, the emotional state is affected and can be changed to another state. For example, in a chat environment, one enters the conversation in sadness state and may end up leaving the conversation happily. This could be due to good news, or high entertainment offered by the conversation etc. The following is postulated based on observations in real life situations:

Firstly, people resist changes to their emotional states. People do not instantly 'flip' from one state to another, and there is a 'momentum' involved in their existing emotions, which tends to be maintained and is referred to in this project by the term 'Emotional Momentum'. This can manifest itself not only by the 'resistance time', which is the time taken for someone to switch from an initial state to a final state, but also by the intensity of the emotional state. One example of the resistance time is that the time taken for someone in a state of sadness to switch to the state of happiness under the influence of jokes.

Secondly, the value of the emotional momentum is proportional to the intensity of the emotional state, e.g. "I am very happy" has higher momentum than simply "I am happy". Therefore, emotion momentum may be displayed by the appropriate expression intensity over a finite period of time. The 'very angry' emotion can be displayed by an expression of a very angry face, but 'very angry' with high momentum emotion can be represented by an angry expression over a relatively long time, or a very angry expression over a shorter time duration. Similar statements can be made for the momentum of other expressions.

Thirdly, the apparent communication of emotional momentum is affected by interaction with people. The momentum has a sign, which can be positive or negative. To assign the signs of emotions, OCC model was evaluated. The OCC model assigns 11 categories to be positive and 11 to be negative [Ortony et al. 1988]. All 11 positive OCC categories are mapped to the happiness expression [Bartneck 2002 and Ekman 1985] while the 11 negative OCC categories are mapped to the four negative expressions of anger, sadness, disgust and fear [Bartneck 2002]. It is argued here that the OCC model is too complex to implement in Internet communication and another problem with the OCC model is that facial expression of surprise cannot be linked to any OCC categories. For the system here it is assumed that the positive emotions are happiness and surprise, and the negative emotions are sadness, fear, anger, and disgust.

Fourthly, personal emotions eventually die out or 'decay' in time but the rate of decay is affected by both positive and negative interactions with other people and by sharing of emotions. For example, one in sad mood can become happier later in the presence of happy people and may become even happier because of support from other people. Another example is that the intensity of someone's sadness may decrease faster when sympathetic feedback is received by others and someone's happiness may make a number of people happy when shared. One can alter a certain emotion momentum by very strong emotional interference over a short time or by sustained lower intensity emotion interference over a longer duration. The emotional intensity depends on the gravitas or perceived reason for being in this emotional state, e.g., sadness due

to death of a friend has normally higher intensity and momentum than sadness due to loss of a material item. The duration of the emotional state depends on individual character culture and generally is also influenced by the frequency of the emotional news. A higher momentum emotion is likely to last longer.

Fifthly, when we read online articles or chat with other online, we may encounter a succession of equal emotional content which interacts with the reader. The most recently read emotional text has the largest influence (momentum) and the previous ones are weaker due to the natural decay of emotional intensity. When dealing with text, we can deduce the emotional momentum from emotional words by using textual analysis methods to identify the emotions and their intensities. The influence may be derived from the history of emotional sentences. In the developed prototype system, the history of emotional sentences is measured by the average weighted emotional mood indicator (AWEMI) of the person, which will be discussed in later sections.

Finally, we introduce the concept of 'personal emotional space' which refers to the set of all aspects, people, and situations that can influence the existing emotional state. The emotional distance or range of influence on others depends on their understanding, knowledge, personal character, common cultural values, common experiences and communications. This is unique to each individual and varies from person to person. For example, you may become happier on the news of a success of someone in your family than a similar success of someone you have never met. The same influence invokes different intensity emotions due to different emotional space. Persons we have never met normally do not belong to our own personal emotional space, and therefore the same news about them does not change our individual emotional momentum.

3.3 The proposed system architecture

The expressive communication system was developed based on the emotion momentum theory. The system includes two components: i.e., the emotion extraction engine and expressive image generator. The engine is capable of analysing textual Internet messages, sensing the embedded emotions and displaying corresponding expressive images (depending on the environment the engine is applied into). The expressive image generator will synthesise facial expression images from a single neutral facial image and the synthesised facial expression images will be used by the emotion extraction engine.

A user may prefer the expressive images drawn by a professional artist or prefer using their own photograph. It is possible that the quality of a professionally drawn image would be much higher than an image synthesised by an expressive image generator. However the expressive image generator can generate facial expressions for millions of potential Internet user and it is hard to imagine that an artist can cope with this amount of potential users. Hence, the automatic facial expression synthesis system – the expressive image generator is provided to create facial expressive images with different intensities for each user.

As the emotion extraction engine was primarily designed for the real-time Internet communication, an Internet chat scenario is presented here to illustrate the working flow of the emotion extraction engine and the expressive image generator: A user logs onto a web site containing the expressive communication system and starts to chat by typing in a text message. Instead of sending the message straight away, the emotion extraction engine will analyse the input message and sense the embedded emotional feelings. When emotions are detected by the engine, corresponding expressive images generated by expressive image generator will be shown to everyone in the same chat room. The workflow of the prototype system is presented in figure 12. The emotion extraction engine and expressive image generator will be discussed from chapter 4 to chapter 6 in details.

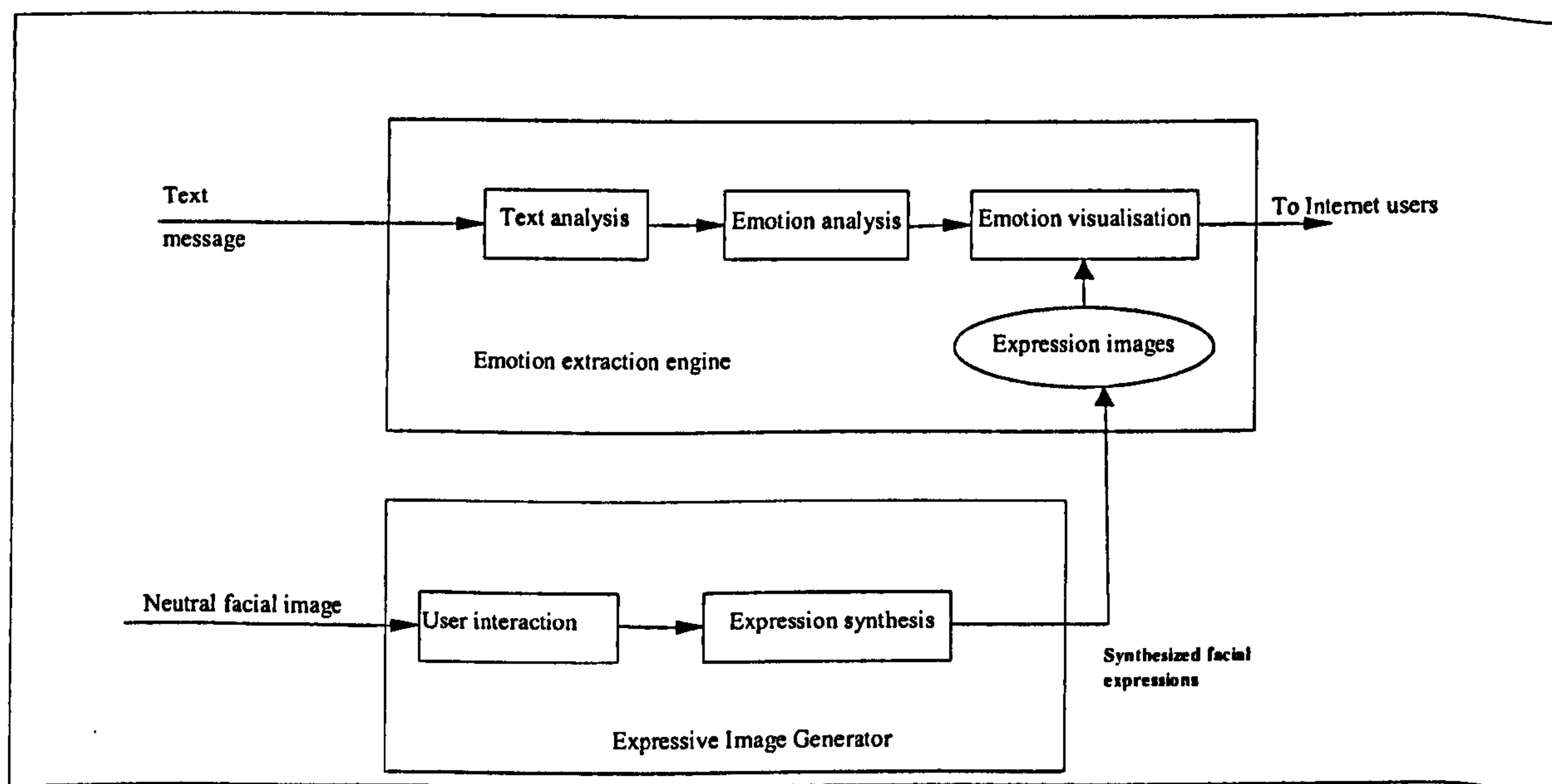


Figure 12: The workflow of the prototype system

In the emotion extraction engine, the keyword tagging approach is adopted as the fundamental analysis method based on two reasons. First, keyword tagging is a relatively fast approach and

can be easily understood by humans. The second reason is that other approaches, e.g. lexical affinity, real-world knowledge, and fuzzy logic analysis, can be combined with keyword tagging in a relatively easy manner. For example in the emotion extraction engine, the keyword tagging method is combined with the real-world knowledge analysis and fuzzy logic analysis to analyse emotions.

When we look back into the history, facial expression synthesis for human images or cartoon images is not a new term. Various algorithms and framework already exist. For example, in Cassell's book "Embodied conversational agents" [Cassell et al. 2000], and in Trappl's book "Emotions in Humans and Artifacts" [Trappl et al. 2003], various computer-generated cartoon-like characters that demonstrate many of the same properties as humans in face-to-face conversation are discussed and various social and emotion factors have been evaluated. One particular interesting work is Metaface framework [Beard et al. 1999], in which a series of cartoon expressions were generated. However, when applied in the Internet communication context, most of the above works suffer from following two aspects. First is the diversity issue - the characters are pre-selected or pre-rendered by artists or 3D packages, hence it is difficult for the vast majority of Internet users to create a personalised character. Second is the computation time issue.

To apply 3D facial models to individual users, it is usual to map the models onto a photograph selected by a user and users need to manually correlate the image to the facial model. This normally requires identifying the positions of more than twenty polygons, which may lead to long computation time and poor user interface interactions.

Conventional 2D facial image generation technologies work in a similar manner to 3D techniques. Facial models are generated from huge facial image databases by extracting and averaging the manually detected features. However, 2D approaches are attractive because that they are relatively free from computational complexity and are supported by some psychophysical theories [Bulthoff and Edelman 1992 and Riesenhuber and Poggio 2000]. It is relatively easy to detect the edge points used in facial model by using various algorithms, e.g. the active appearance models [Cootes et al. 1999]. Although the performance is arguable, these automatic edge extraction algorithms can considerably relieve the painful user interaction process of facial generation and may be a leading algorithm in the future. For real-time Internet communication, fast and user-friendly interfaces are most desired. Hence 2D approaches have been adopted for the expressive image generator.

The expressive image generator uses image warping and morphing approaches to synthesise facial expressions. From the background literature, it can be seen that image morphing and warping had been used to generate facial distortions by former researchers. By combining image warping and morphing, recognisable facial expressions may be generated.

3.4 Chapter conclusion

This chapter first discussed the emotion theories that were followed, and the emotion momentum theory that was developed for this project. The emotion momentum theory treats emotion as a three dimensional concept – the emotion category, the intensity and the duration, all of which will influence the perception of emotions. The emotion momentum theory was implemented in a prototype system which contains an emotion extraction engine and an expressive image generator. The emotion extraction engine can analyse textual messages, detect embedded emotions and visualise emotions via expressive images generated by the expressive image generator. From next chapter, the two components in the prototype system – the emotion extraction engine and the expressive image generator will be discussed in detail. Chapter 4 will focus on the system architecture of the emotion extraction engine and the assessment of its performances (i.e., speed and accuracy). Chapter 5 will focus on an alternative technique – fuzzy logic, which is applied in the emotion extraction engine to analyse conflicting emotions and mood.

Chapter 4

The emotion extraction engine

Last chapter describes the emotion momentum theory and the proposed framework to detect and visualise emotion for Internet communication. In this chapter, the emotion extraction engine, which detects emotional feelings from textual messages, is presented in detail. First, the requirements of the emotion extraction engine are discussed, and then the architecture and individual components of the emotion extraction engine are examined. A series of experiments assessing the performance of the emotion extraction engine, i.e., speed and accuracy, are described as well. In the final part of this chapter, an experiment assessing the engine's ability of handling conflicting emotions is presented.

4.1 The requirements of the emotion extraction engine

To develop the emotion extraction engine successfully, the requirements of the engine, e.g. interface requirements and communication requirements etc, should be clearly defined. Various related web sites and products were observed, e.g. yahoo chat rooms and MSN messenger, and hold several discussions with potential internal users, i.e. colleagues, and external users, i.e. students.

The emotion extraction engine is mainly designed for real-time Internet chat environment. The emotion extraction engine should be able to analyse the textual message typed by the user in real-time in order to sense the emotional feelings contained and present these feelings to other users. The emotion extraction engine may also be applied in Internet game environments, e.g. online game etc. Similar to the chat environment, the emotional feelings may be detected from users' textual messages. The difference is that in an online game environment, emotional feelings may occur when the player achieved a specific task in the game, e.g. just conquered a castle or killed a very tough opponent etc.

Another environment the emotion extraction engine may be applied in is the article analysis environment. For example, a user may input an article to the emotion extraction engine in order to find out the overall emotion of the article. The input for this scenario may contain numerous

sentences and the sentences' structures would be much more complex than the Internet chat messages.

4.1.1 Interface requirements

The interface requirements of the emotion extraction are different between the online environments (e.g. online chat and online game) and the offline environment (e.g., article analysis environments). The requirements for Internet chat and online game environments include the provision of facilities that enable users to communicate with each other via text and voice. The detected emotion of each user should be visualised as facial expressions of humans or cartoons. The interface should provide a facility to remember each user's emotion history (e.g. if a user has been happy during the previous ten chat messages) as an indicator and the indicator will be able to be viewed by every user. The process of connecting to other computers on the network should be as simple as possible, e.g. users should only need to indicate the person they wish to connect to rather than enter IP addresses or other information. The system should encourage interaction between users and whoever is able to communicate with through the network.

Interface requirements for the article analysis environment include the provision of facilities that enable users to input articles. Users should be able to either copy-paste an article or load an article directly to the emotion extraction engine. Indicators should be provided for each emotion category that the emotion extraction engine can detect. The detected emotion of each category should be totalled and displayed visually in the indicators to provide a general picture of the relative differences between the different categories of emotions sensed in the article. A facility to calculate and display the overall mood of the article should be provided as well.

4.1.2 Communications requirements

Online environments and off line environment have different communication requirements for the emotion extraction engine. The chat environment and online game environment require a real-time emotion extraction engine as users' textual messages are typed quickly and the delay triggered by the emotion extraction engine should be kept to a minimum. As the Internet speed of most home users is still limited, the emotion extraction engine should transmit small amount of data, which would be just enough to indicate which corresponding expressive images to display.

For the online environments, the emotion extraction engine should provide users with facilities to communicate via text, which requires following facilities:

- The facility to send text messages directly to other computers within the emotion extraction engine system using TCP/IP.
- The facility to send parameters to users to display expressive images using TCP/IP.
- The facility to generate and display expressive images for every user.

The article analysis environment does not have an urgent real-time requirement as the online game or chat environment does. The textual messages are contained in an article and are typed or copied in an asynchronous way. However the delay triggered by the emotion extraction system should be kept to a minimum as well. As the hardware acceleration continues, computers are utilising more memory and faster processors. This enables a more complicated structure for the emotion extraction engine compared to the engine used on online.

4.1.3 Input and output requirements

The interaction between the user and the system should be simple and users should be able to access the features with the minimum of effort. The number of steps required to operate the different features of the system should be kept to a minimum. All supporting functions should be running in the background and provide instant access to all services the users require.

Input into the system should involve intensive use of the keyboard and mouse. Output by the system to the user should be clear and attractive using multimedia formats without being text intensive. The system should provide a text-to-speech option to read out any text that is displayed. The presentation is to be imaginative, friendly and easily comprehended. Users who do not have experience with computers and those who are disadvantaged by age or handicap should be able to use the system intuitively.

4.1.4 The requirements of the development language

Most online chat environments are based on web browsers. This restricts the application size, user interface design and users' interaction methods. However products like MSN and Yahoo Messenger are standalone software that enables real-time communication. Currently the most popular programming languages for chat environment development include ASP, PHP, JAVA, Visual Basic and JavaScript. To analyse and extract emotion contained in a textual message, database programming and complex logic analysis are inevitable. JAVA and Visual Basic are more appropriate for this project as they can be used not only for network browsers, but also for

stand alone products. To reduce the learning curve and focus on research problems, Visual Basic was chosen to be the main development language for the emotion extraction engine.

4.1.5 Requirements of the performance assessment of the emotion extraction engine

The most important factors in assessing the performance of the emotion extraction engine are the engine speed and the accuracy. The engine speed and the accuracy of engine are both crucial from the point view of users. If the engine needs more than a minute to analyse one sentence that is typed in, then from Human-Computer Interaction (HCI) point of view, it can not be suitable for the purpose of Internet communication. On the other hand, if the engine always finds "happiness" emotion in a sentence while users believe it should be "sadness", then from technical point, the engine failed.

To assess these two factors, users' textual messages needed to be input into the engine. Because of the scope of this project, users were randomly selected from the students and staff of Bournemouth University. Except for users' real-time input, textual messages were collected from the BBC web site and online chat archives that discussed different issues in order to test the engine with diverse range of text messages from different backgrounds (e.g. an archive focusing on gossips about celebrities and an archive of discussions about computer science issues).

In order to determine whether the accuracy of the engine is significant or just by chance, statistical analysis were applied to the data generated by monitoring the performance of the engine. T-tests were used to assess whether means of two groups of data were statistically different from each other. To assess the links between emotions embedded in articles and the article contents, correlation tests were used to assess the trend between two sets of data.

4.2 Emotion extraction engine system architecture overview

With the significant hardware improvements in recent years, users' desktop PCs are powerful enough to carry out scientific analysis that used to be done in mainframe machines. Thus, instead of placing the emotion detection workloads on a dedicated server for every user, the workload is distributed to user's machine. This means that the emotion extraction engine is installed at client side. Figure 13 presents an overview of the emotion extraction engine.

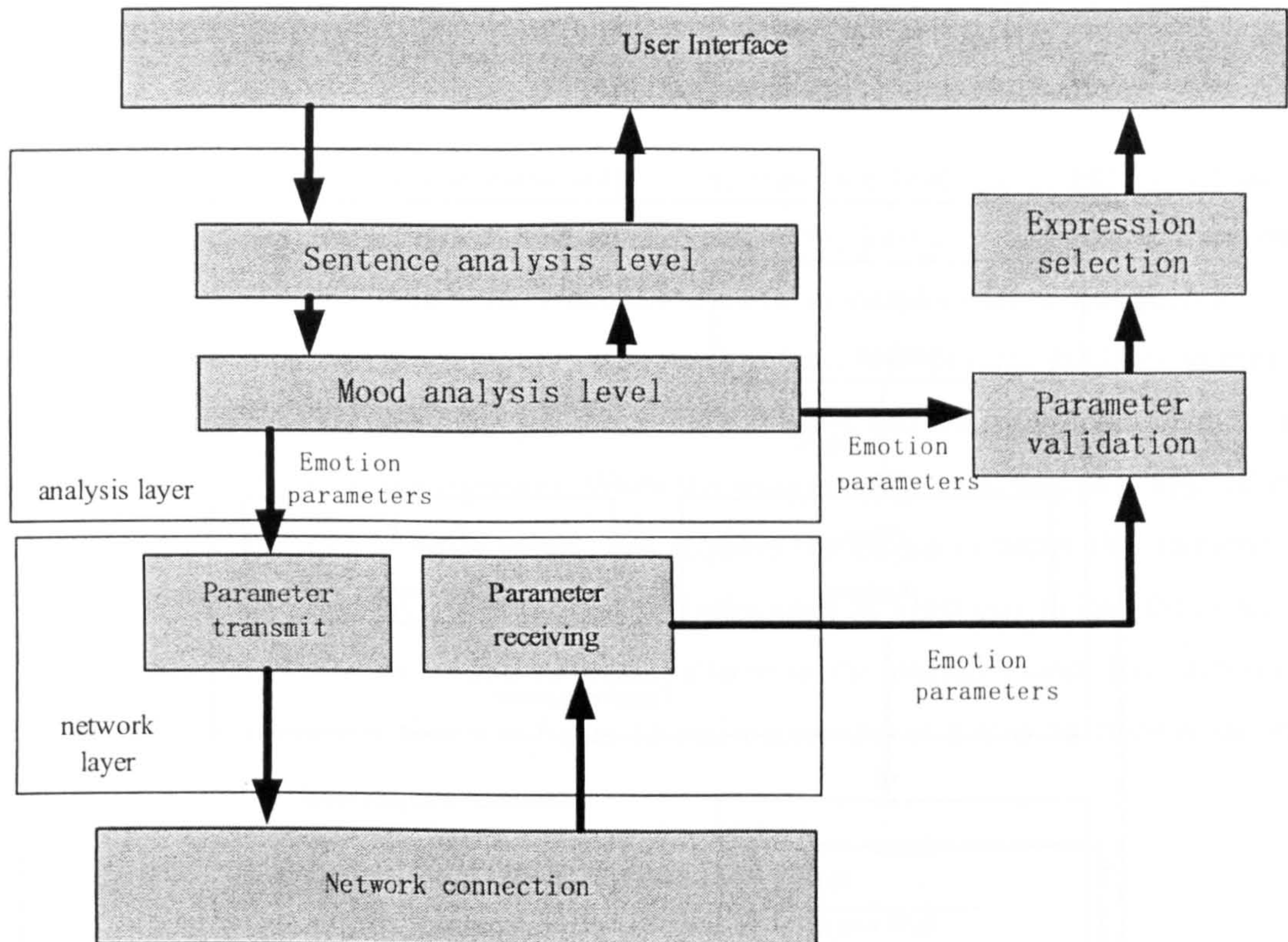


Figure 13: Overview of the emotion extraction engine

The components of the emotion extraction engine are classified into two layers: the *user interface layer* and *analysis layer*. The *user interface layer* is in charge of receiving textual messages from local user and display expressive images corresponding to the emotion detection analysis result. The *analysis layer* detects the emotional feelings presented by the local user. A set of parameters to trigger the expressive image display is generated in this layer. Other components of the emotion extraction engine include the *parameter validation component*, *expression selection component* and *network layer*. The *parameter validation component* undertakes the responsibility of validating the received parameters (e.g. validate parameter format and value); the *network layer* sends and receives the parameters and users' textual messages. When emotions are detected in the *analysis layer*, the generated parameters will be sent to the *expression selection component* and the *network layer*, which in turn selects the expressive image to be presented and sends the parameters to the network. When the parameters generated by other users are received in the *network layer*, the parameters will be delivered to the *expression selection component* and the corresponding expressive images will then be displayed.

The analysis layer can be classified into two levels: the *sentence analysis level* and the *mood analysis level*. The sentence analysis level detects the emotional feelings embedded in a single sentence, while the mood analysis level stores and analyses the result from the sentence analysis

level to discover the current mood and resolve possible conflicting emotions. The structure of the analysis layer is presented in figure 14.

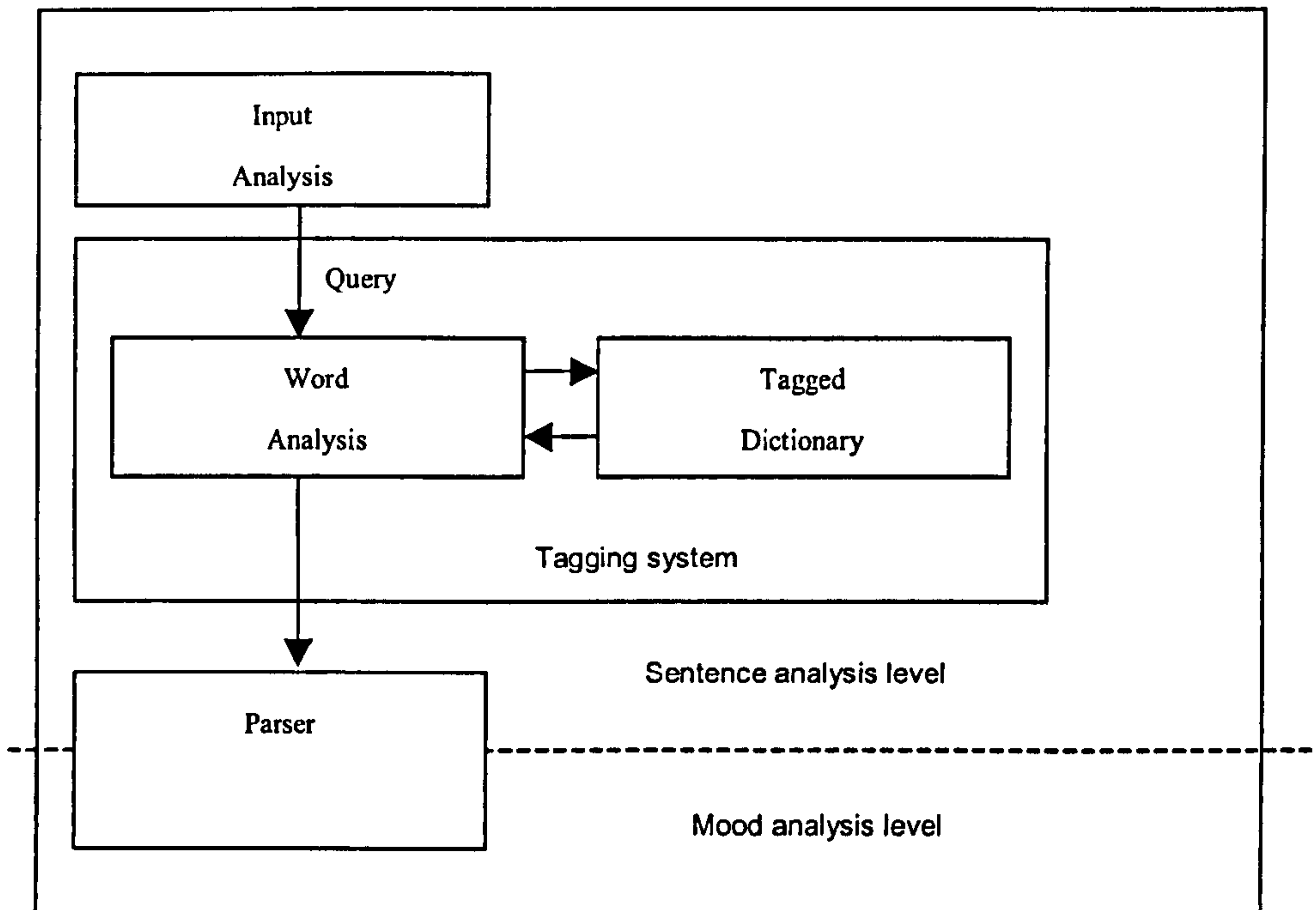


Figure 14: Structure overview of the analysis layer

In the analysis layer, the *input analysis function*, the *tagging system* and parts of the *parser* component belong to the sentence analysis level. The mood analysis level contains the mood and conflicting emotion detection components of the *parser*.

The input analysis function divides the textual messages into words and punctuation; the tagging system converts words into their word categories by searching the tagged dictionary; in the sentence analysis level, the parser applies rewrite rule [Russell and Novig 2002], real-world knowledge and linguistic knowledge to detect the possible emotional feelings; in the mood analysis level, the parser applies history information and other techniques including mood average and fuzzy logic to acquire the current mood. In this chapter, the emotion average technique is discussed while the fuzzy logic method will be presented in next chapter.

To guide the emotion extraction engine's analysis, 26 rules have been deducted from linguistic knowledge, real-world knowledge and fuzzy logic. The classic rule based architecture was applied in the emotion extraction to direct the analysis process and the output data was organised in semantic way. The following sections will focus on the layers of the emotion extraction engine and rules applied on each.

4.3 The interface layer

Besides interacting with users, the main duty of the interface layer is to display the current emotion status and the current mood. The appearance of the interface depends on the context that the emotion extraction engine is in. When the emotion extraction engine is applied in a real-time Internet communication environment, the interface layer will provide facilities to transmit chat messages, display expressive images for individual users and manage screen layouts, e.g., font management and colour management. When the emotion extraction engine is applied in an article analysis context, i.e., an environment that requires the engine to detect the emotions of a specific article, the interface layer will provide facilities such as a text box to load the article, an indicator to show the current mood and an indicator to show the average mood. The snap shot of an article analysis interface is shown in figure 15 and a snap shot of a chat interface is shown in figure 16.

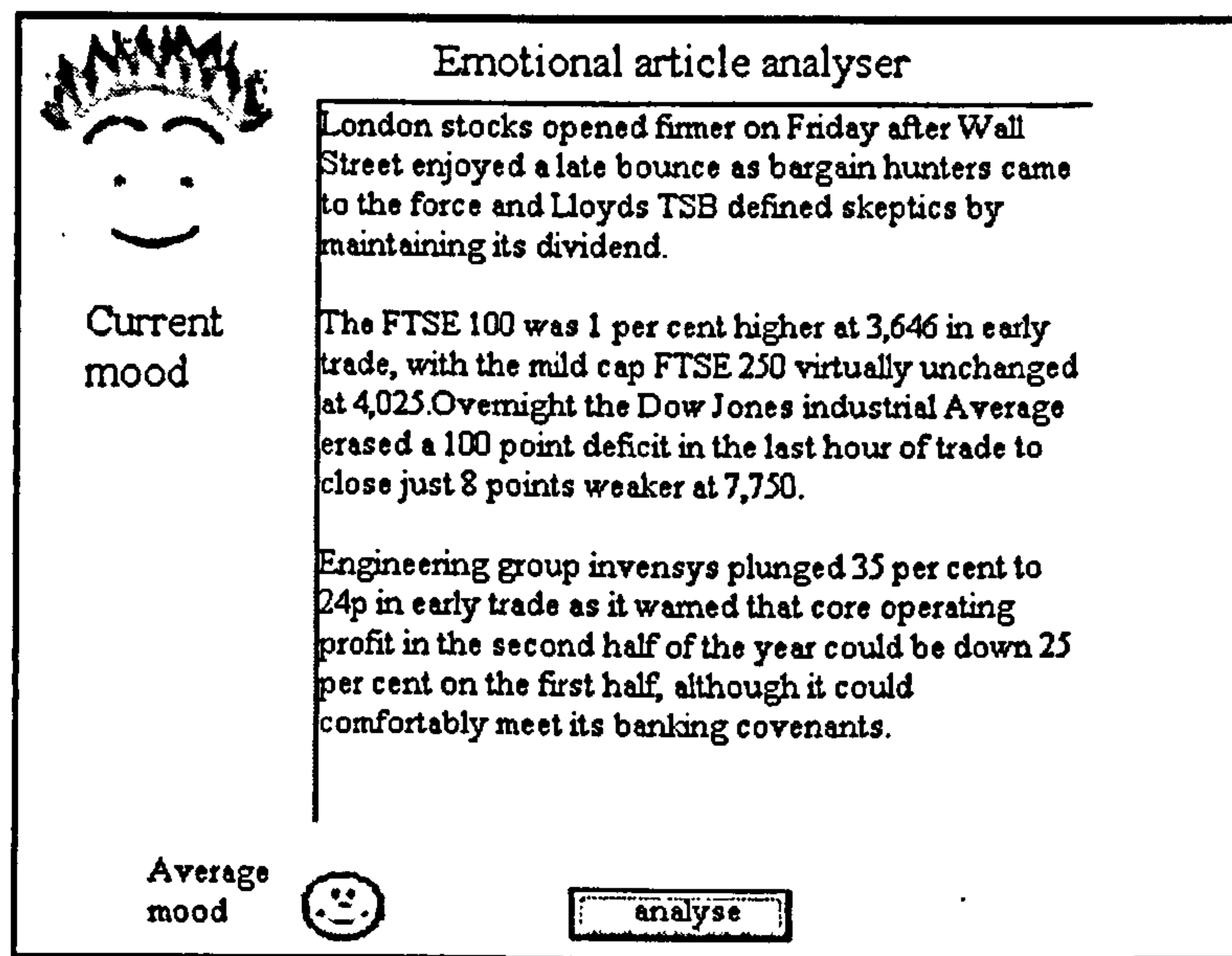


Figure 15: Article analysis interface

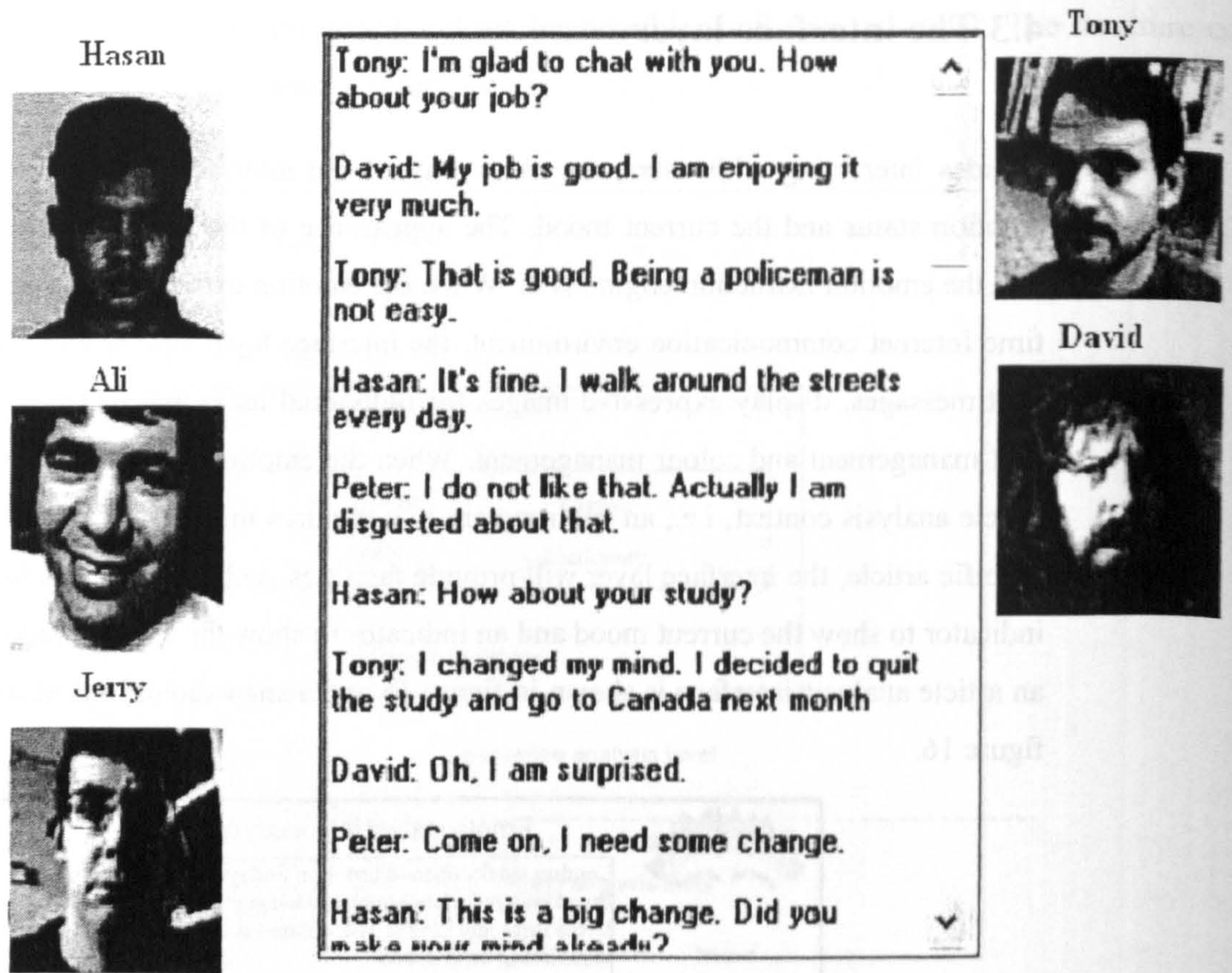


Figure 16: Expressive 2D chat interface

The expressive images are displayed in succession based on the succession of the emotions detected in the sentences. The maximum duration of an expressive image being held on screen is determined the AWEMI, which will be discussed in rule 25. However, if the emotions alter at a faster rate, then they are displayed accordingly.

4.4 Analysis layer

The analysis layer is divided into three sub components: input analysis, tagging system and parser (see figure 14). In this section, these three components are in turn discussed. Keyword tagging and real-world knowledge was chosen as the fundamental sentence analysis mechanisms for the emotion extraction engine. The speed of keyword tagging is fast, and it easily detects the direct meanings. However pure keyword tagging cannot handle the real-world situation since the semantic meanings is not available. As a result, real-world knowledge is taken into account in the analysis.

4.4.1 Input analysis

As mentioned before, the input analysis function divides the textual messages into words and punctuation. Users' input sentences are sent to the input analysis function for initial assessment. The length of a user's input can vary from just a few words to a-hundred-thousand-word article. An article may contain hundreds of sentences, and even in a chat environment, users may type in once more than one sentence. The output from the input analysis will be passed on to the tagging system. To simplify the tagging system process, sentences are separated by the input analysis function, so the tagging system only handles one sentence a time.

There are some special cases that the input analysis function needs to deal with. For example, a dot can be not only a terminator, but also a decimal point. Decimal points are widely used in published articles and chat environment. Another example is exclamation mark and question mark. Exclamation marks and question marks terminate a sentence in most cases, however in sentences such as "I am really good !!! or what ???", the first exclamation mark or question mark does not terminate the whole sentence, instead, these marks represent an increase of emotion intensity. The third example is that in chat environments, users tend to omit the termination mark. When the input analysis function faces a sentence without a termination mark, a full stop should be added. Rule 1 presents the guidelines of the input analysis function.

Rule 1. Sentence terminations

To correctly terminate a sentence, the input analysis function will carry out following analysis steps: Firstly phrases such as "Mr.", "Mrs.", "," and ":" will be replaced by pre-defined symbols. Secondly when marks such as "!", "." or "?" are detected, the emotion extraction engine will continue searching the repeating marks to detect sentences such as "I am really happy !!!" or "I mean". Thirdly, when no termination marks are found in a sentence, a full stop symbol will be added. The input analysis function generates individual sentences as output and sends those sentences to the tagging system. The procedure of the input analysis function is illustrated in figure 17.

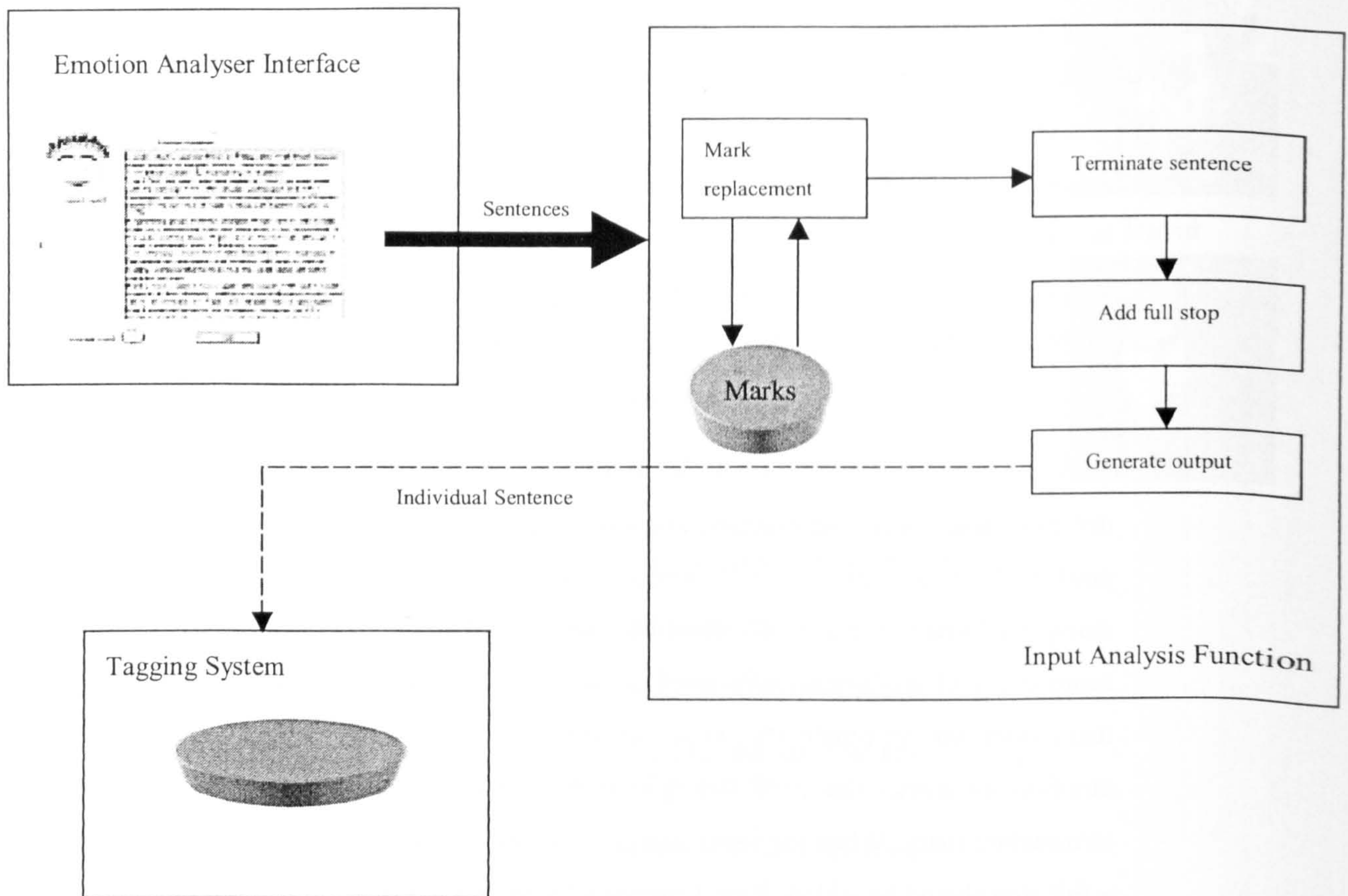


Figure 17: The input analysis function

4.4.2 The tagging system

The tagging system contains words and their correspondent categories and provides the fundamental knowledge for the parser of the emotion extraction engine. A tagged dictionary and word analysis function constitute the tagging system.

4.4.2.1 Tagged dictionary

To create the tagged dictionary, standardised word corpora, e.g. Brown corpora [Francis and Kucera 1979] and BNC [Leech 1997], were examined. Instead of using a general purpose corpus, a dictionary of over 18,000 words was created solely for this project. The entries in the dictionary contain the words, tags indicating which emotion category they belong. To find words with possible emotional feelings, dialogues in textual chat environments, English dictionaries and the BBC website were manually searched through. The dictionary can be tailored to fit in different contexts, e.g. stock market articles. In order to keep a minimum response time, the tagging system requires the entire word in the dictionary to be tagged.

The dictionary is implemented as a Microsoft Access database and is connected to the analysis functions with the ODBC programming. The main table "word-tag" in the database contains information about words and the tags, which in turn consists of three fields, i.e., word, category and detailed category. The field "detailed category" is always null unless the value of the field "category" equals "EMO_W", which indicates an emotional word. When the field "category" is EMO_W, the value of the field "detail category" represents the emotion category, word type and emotion intensity, e.g. 1N2. To prepare and tag the dictionary, following rules – rule 2 to rule 4, are defined.

Rule 2. Explicit and inferred emotional words

Emotional words are divided into two categories: the inferred emotion category and the explicit emotion category. Explicit emotional words directly convey an emotion while inferred emotional words do not directly convey an emotion but describe situations that contain emotional feelings. The intensities of inferred emotional words are lower compared to explicit emotional words. For example, it is reasonable to assume that people are happy when they buy something. The sentence "I bought a car" may present the emotion happiness, however the emotion intensity is lower than "I am happy".

Rule 3. Word ambiguity

Even a unique tag is specified for each word group, there are words (e.g. pretty) that may appear in several groups and therefore have to be tagged to accommodate for multi-possibilities. For example the sentence "I got a pretty car" presents a happy feeling while the sentence "I got a pretty ugly car" may presents a feeling with dissatisfaction. To deal with this problem, a special tag "AM_EMOTION", which stands for ambiguous emotion words, is designated.

Rule 4. Address words

Some emotional words are used in phrases to present locations, addresses, names and some traditional meanings. In this case, there are no the emotional feelings with those words. For example, "new" is a word with inferred emotion, while the phrase "New York" does not contain any emotional meanings on its own. A collection of addresses and names are saved in the dictionary to distinguish the address words from any emotional words.

4.4.2.2 Tag set

Tag set is the universe of tagging categories. Although different tagging methods have been proposed by numerous researchers [Leech 1997 and Francis and Kucera 1979], these sets do not

fulfil the engine's requirement as conventional tagging methods only specify the suffixes or prefixes required to differentiate between different word groups. The emotion extraction engine requires the entire word to be tagged in order to improve the speed. For every emotion word and its' related intensity, the engine requires particular marks to represent its emotion category and intensity. A tag set, which includes 119 categories, was solely developed for the engine. A brief description of a subset of the tags is presented in table 1.

Tag	Description	Example
NOUN	Noun	kindness
1N2 EMO_W	Happiness, Noun, intensity 2	happy
2N2 EMO_W	Sadness, Noun, intensity 2	sad
3N2 EMO_W	Fear, Noun, intensity 2	fear
4N2 EMO_W	Surprise, Noun, intensity 2	surprise
5N2 EMO_W	Anger, Noun, intensity 2	angry
6N2 EMO_W	Disgust, Noun, intensity 2	disgust
6N3 EMO_W	Disgust, Noun, intensity 3	aversion
PCF	Present continuous	am watching
PP	Present perfect	have seen
PAST	Past	was
VI	Verb	abandon
NT	Noun, third person	she
AMEMO_W	Emotion with ambiguity	pretty

Table 1: The tag subset

In the tag set, 93 tags (e.g. tag Noun and tag Verb) are classified as ordinary non-emotion tags and another twenty six are designated to represent emotion. In table 1, the tag "EMO_W" indicates the explicit emotional words, and the tag "INFERRED" indicates emotional words relating to an inferred emotion state. In the tag set, emotional words are further classified into six emotion categories: *happiness*, *sadness*, *surprise*, *fear*, *anger*, and *disgust* with three different emotion intensities: *low*, *medium* and *high*. Accordingly, the tag set uses numbers from 1 to 6 to represent the emotions from *happiness* to *disgust*. Each emotion word tag also provides the intensity information with numbers. For example in table 1, tag *1N2* represents a noun in emotion category *happiness* with a medium intensity.

In the tag set, a special tag *NDP* (negative data point) is assigned to the words that have opposite influences on the emotion information. For example, in the sentence "The man halted their happiness", the word "halted" will overturn the emotion feeling of the sentence and therefore "halted" is assigned as *NDP*. Another special tag *DEDUCTION* (deduction word) is assigned to the words that have influences on the emotional information that can be removed from the sentence. For example, in the sentence "I would rather be happy, instead of being sad", word "instead" will totally remove the sad feeling, therefore the tag of word "instead" is assigned as *DEDUCTION*.

Some words can increase or decrease the intensity of emotional feelings. For example, the word "very" in sentences such as "I am very happy!" and word "little" in sentences such as "I am a little sad" both influence the amount of emotional feelings presented. Tags I (increase) and D (decrease) are assigned to these types of words. As the influence on the emotion intensity varies with different words, the tag-set uses the numbers from 1 to 3 to represent the different amounts of influence on intensity (e.g. a word belonging to tag I2 will increase emotion intensity more than a word in tag I1).

4.4.2.3 Word analysis component

When receiving the output sentence from the input analysis function, the first step of the tagging system is to split the sentence into individual words. Then the word analysis component will check through the tagged dictionary to look for each word and the corresponding tag category. If a word is not found in the dictionary, the engine will undergo a suffix and prefix analysis. By examining the suffix and prefix of the word, the assumed tag may be derived; otherwise tag "unknown" will be assigned.

Some verb phrases or noun phrases have meanings equal to "not", e.g., "I will give up yesterday's gains". The phrasal verb "give up" represents a negative feeling and the sentence is equal to "I will not gain". Some phrases may erase emotional feelings, e.g., "Rather than sad, I am actually quite surprised". The phrase "rather than" erases the sadness emotion totally. The word analysis model searches for phrases with possible negative meanings and erasing functions. When finding these phrases, a special mark will be assigned.

If the word's tag is "EMO_W", which hints that the emotional feelings are contained in this word, the analysis function will pass its position and tag details to the parser component for further analysis. The tagging system provides the foundation for the parser component by providing vital information about words and corresponding word-category tags. The output of

the tagging system, i.e., words and corresponding tags, are input into the parser component to analyse the possible embedded emotional feelings. The workflow of the tagging system is shown in figure 18.

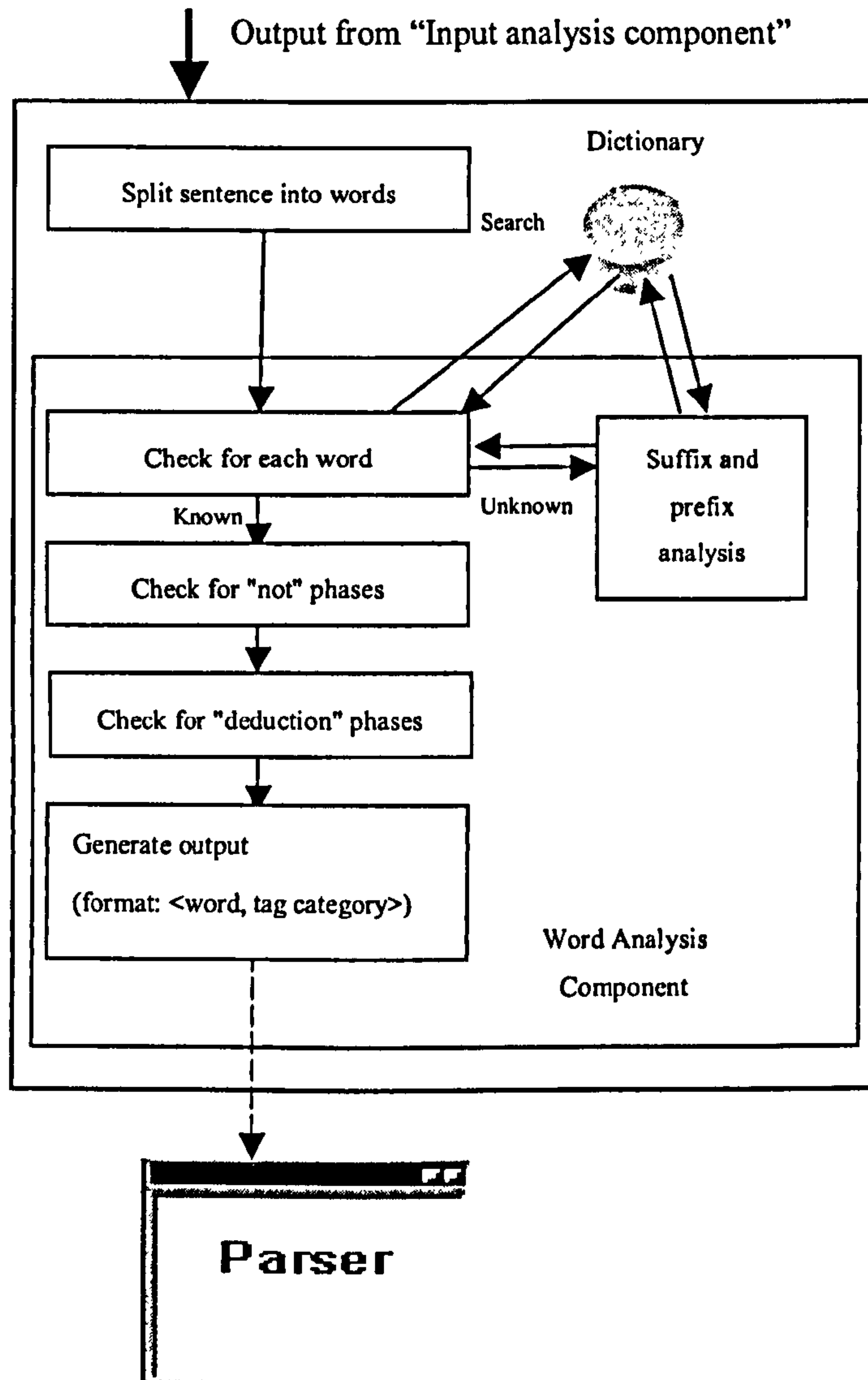


Figure 18: Tagging system workflow

4.4.3 The parser

The parser receives the output from the tagging system and the initial parsing procedure is accomplished through the use of rewrite rules [Russell and Novig 2002]. The parser then undertakes two levels of parsing functions. The first level is individual sentence parsing and the second level is the mood analysis and the detection of conflicting emotions. Two different approaches are implemented for the mood analysis and conflict emotion detection level. One is

the mood average method, which will be discussed in following sections; the other is the fuzzy emotion analysis, which will be described in chapter 5.

4.4.3.1 Individual sentence parsing

When an emotion word is found in the initial parsing procedure, emotion extraction engine rules 5 to 19 will be applied to guide the sentence parsing. When no emotional words are sensed, rules 20 to 24 will be utilised.

Rule 5. Emotion symbols

In a chat environment, some emotion symbols (e.g.:-), :- () are widely used to represent emotional feelings. The engine will search for these emotion symbols and map them to corresponding emotional words, e.g., symbol :-) is mapped to the word happy. Most emotion symbols are only valid in the Internet communication environments, thus when the emotion extraction engine is applied into other environments, this rule should be ignored.

Rule 6. Emotion acronyms

In a chat environment, acronyms, e.g., LOL (laughing out loud) and LMAO (laughing my ankles off), are widely used to represent emotions with high intensities. However these acronyms have almost never been used outside the chat context and the life span of their usage is limited. As a result, these acronyms are stored in the dictionary but are only checked when the parser is applied in a chat environment.

Rule 7. Acronyms and emotion symbols in chat environment

In a chat environment, acronyms and emotion symbols strongly dominate the emotional feelings. For example the sentence "You are ugly :-)" presents an ironic and happy feeling despite the existence of the emotion word "ugly". When emotion acronyms or symbols are found, the parser will set the sentence's emotion according to the category of the acronym or symbol and discard other emotional words.

Rule 8. Positive emotions and negative emotions

The engine classifies emotional words into six categories: *happiness*, *sadness*, *anger*, *fear*, *surprise* and *disgust*. However, some emotional sentences may present feelings contained in more than one category. For example, the sentence "my car was crashed" contains emotional words "crash" which belongs to the sadness category according to the tagging system. However,

this sentence may be perceived differently by different human viewers, e.g., as a sentence combining feelings such as sadness, fear or anger.

To solve the problem, two more general categories were introduced: positive emotion category and negative emotion category. The positive emotion category includes emotion categories happiness and surprise and the negative emotion category include emotions categories sadness, fear, anger and disgust. The sentence "my car was crashed" then can be classified as a negative emotional sentence.

Rule 9. Recorded information of emotional words

When emotional words are found in a sentence, the parser will search for the referred participant of the emotional words, the tense of the sentence and the conditions under which the emotion occurred. Table 2 illustrates the examples of recorded information.

Sentences	Participant	Tense	Condition
I am happy.	I	Present	Definite
I think he is happy.	He	Present	Definite
I will be happy if he comes.	I	Future	Conditional

Table 2: Example of recorded information of emotional words

Rule 10. Conditional emotions

Conditional emotions present emotional feelings that may or may not happen. The parser considers a conditional emotional sentence as an emotional sentence with reduced intensity. For example, if an sentence such as "I will be happy" is assigned as the emotion category of happiness with an intensity of 2 because of keyword "happy", the sentence "I will be happy if he comes" will be assigned as happy with an intensity of 1.

Rule 11. Multiple emotional words

If there is more than one emotion word in a sentence and they are connected by a conjunction then the parser will treat it as a sentence with multi-emotion states, e.g., *'I am surprised and happy about it'* will be treated as a sentence with emotions of surprise and happiness. This rule will not combine or deduct any emotions. Multiple emotions detected by this rule will be analysed by dedicated conflicting emotion detection methods, which will be described in later sections.

Rule 12. Null participant

If no participant is found during the sentence parsing, the engine will refer the participant to the person who is in communication or the default subject. When this rule is applied in real-time communication environment, sentence such as '*apprehensively happy*' will be treated as '*I am apprehensively happy*'. When it is applied in the stock market article analysis environment, sentence such as '*extremely happy*' will be treated as '*the market is extremely happy*'.

Rule 13. Phrase checking

Some words may contain no emotion meanings themselves. However, when they are constituted into phrases, e.g. "cold feet", emotional feelings will emerge. As these phrases rely on daily usage and normally can not be conducted by grammatical rules or semantic rules, the phrase checking will highly depend on the context, i.e., different culture and usage scenarios that the emotion extraction engine applied in.

Rule 14. Emotion intensity

If no adjectives (expressive adjectives) are detected in front of the emotion word then the intensity will depend only on the word's tag category. If there is only one adjective then the parser will increase or decrease the intensity automatically. When more than one adjective is detected, then the parser will increase or decrease the intensity of the emotion depending on the number of the adjectives the parser found, until the emotion reaches the highest or lowest intensity. For example, the parser interprets 'very sorry' as a high intensity emotion phrase while the parser will give sentence "I am less sorry now" a low emotion intensity. For sentence "I am very very happy", which contains a middle intensity emotion word "happy", the parser will increase the intensity of the sentence's emotion to high because of the two adjectives "very".

Rule 15. Sentences beginning with auxiliary verbs

If a sentence begins with an auxiliary verb, the sentence is a question. Emotions embedded in questions present different feelings compared to other sentences (e.g. sentence "Are you happy?" does not represent an emotion related to happy, instead, it provides a sense of puzzlement). The emotion extraction engine is not designed to extract emotion from questions except by using the question mark rule, which will be discussed in rule 22. Thus when sentences begin with auxiliary verbs, they will be checked by rule 22.

Rule 16. Negations

If an emotional word is found in negative form, the parser will discard the emotional feelings in the sentence because the negation of an emotion does not automatically indicate the presence of the opposite emotion, e.g. the sentence "I am not surprised" will be classified as a neutral emotion sentence. It is arguable that sentences such as "I am not happy" may infer a sad or angry feeling. However, as most negative form emotional sentences do not provide an emotion feeling, this rule is correct in most cases. The exceptional sentences can be solved by adopting real-world knowledge, e.g. not happy infers "anger" emotion.

Rule 17. Emotional nouns

Some emotional words will grammatically fall into the noun category and are used as such. The parser will treat these words as normal emotional words. However, when emotional nouns are used as subjects, the emotion extraction engine will treat these words as non-emotional words. For example, "Happiness will benefit your health." will be treated as a non-emotional sentence.

Rule 18. Emotion deductions

In some conditions, emotional words may not represent the intended emotion meanings. For example, the sentence "I was happy rather than sad" presents a happy feeling although the word "sad" refers to a sad feeling. Phrases such as "rather than" removes the emotion feeling that follows. A list of phrases with deduction functionalities is created. As these phrases may remove either the proceeding emotional feelings or following feelings, no general rules are created. Instead, the engine will look for each individual phrase and remove emotional feelings according to individual phrase's usage.

Rule 19. Sentences with conflicting emotions

When more than one emotional words are detected in a sentence it is possible that the emotion meanings contained in different words may conflict. For example, the sentence "I was fine but my car was damaged a little" includes two emotional words "fine" and "damaged". There are no clues within this sentence to indicate whether this sentence presents a happy or sad feeling. To solve this type of problem, two methods: average mood calculation and fuzzy logic analysis were applied. In this chapter, the average mood calculation will be discussed.

Rule 20. First sentence rule

This rule is only valid for Internet communication environments. It is reasonable to assume that most users are polite or have a low intensity happiness emotion at the start of a dialogue. When

no emotional words are detected in the first sentence, a happiness emotion with lowest intensity will be assigned.

Rule 21. Exclamation mark rule

When marks such as "!" or "!!" are detected, it is reasonable to assume that users want to highlight some important issues. Therefore when exclamation marks are found, the intensity of corresponding emotions detected in this sentence is increased. When no emotional words are found, the parser will assign a happiness emotion with lowest intensity to the sentence.

Rule 22. Question mark rule

Sentences that finish with a question mark "?" often contains a puzzlement feeling. In the emotion extraction engine, puzzlement is included within the emotion category "surprise" with the lowest intensity. As a result, the sentences finishing with question mark will be assigned to emotion category surprise with the lowest intensity when no emotional words are found.

Rule 23. Repeated words rule

Even when no emotional words can be detected, some specific repeated words may present inferred emotional feelings, e.g., "no, no, no, you should go left". When these specific words are found consecutively, the engine will assign an inferred emotion to the sentence according to the word category with intensity equal to the number of consecutive words presented.

Rule 24. Neutral emotion rule

If no emotional words can be detected in a sentence and the sentence does not qualify for all above rules, the emotion extraction engine will assign a neutral emotion with medium intensity to it. The middle intensity neutral emotion will be used in the mood analysis and conflicting emotion analysis.

4.4.3.2 Mood analysis and conflicting emotion detection

The emotions contained in each sentence are stored for the analysis of mood calculation. The following emotion engine rules give the principles of mood calculation and conflicting emotion analysis. Before the mood analysis functions are discussed, there are two concepts - "current emotion of a user" and "average mood of a user" that need to be clarified first. The current emotion is based solely on the information presented within a single sentence. In contrast, the average mood is the overall emotion perceived from a series of the user's inputs. The average mood is a calculated result based on the current and previous weighted emotions. Both "average

mood" and "current emotion" contain the emotion category and the intensity. The following two rules (25-26) are defined to calculate the mood and handle the conflicting emotion.

Rule 25. Average weighted mood calculation

The concept of the average weighted emotional mood indicator of the user is introduced, which calculated as follows:

$$\text{Average Weighted Emotional Mood Indicator (AWEMI)} = \frac{1}{N} \sum_{i=1}^N \alpha_i E_i \quad (4.1)$$

where 'i' is an index increasing from 1 to N corresponding to the sentences remembered, or taken into account, up to maximum N. N is chosen to be 5 in this implementation. E_i is the sum of the signed emotion intensities (momentum, with a positive or negative sign) of i^{th} sentence and α_i is the weight given to each emotion within a sentence. E_1 for example, corresponds to the first sentence's total emotion intensity and is the signed sum of all the intensity of the emotions within the sentence (happy, very sad, etc.). Similarly E_2 for the second sentence and subsequent sentences up to the previous five are correspondingly labelled and used in the system. three intensity gradations for each emotion category are adopted, '1' for low level intensity, '2' for medium and '3' for high intensities. In addition, two clusters of emotion categories carrying a positive and a negative sign are defined. The categories 'happiness' and 'surprised' are given a positive sign while the other four (anger, disgust, fear and sadness) are assigned a negative sign.

It is also reasonable to assume that the mood of the last sentence should carry more weight in the overall representation of the mood of the user. Therefore greater weight is given to the emotions contained in the most recent sentence with progressively less weight to earlier sentences: for example $\alpha_1 = 0.5$, $\alpha_2 = 0.4$, $\alpha_3 = 0.3$, $\alpha_4 = 0.2$, $\alpha_5 = 0.1$. Here for simplicity, it is assumed that only emotions in the last five sentences are actively influencing the AWEMI.

The emotion extraction engine will display the appropriate expressive images representing the detected emotions. The engine detects and displays six emotions and stores three intensity levels per emotion, therefore eighteen images are used. The display of the appropriate images is straightforward providing the emotion is clearly identified as one of the six basic emotions. However in many sentences, the emotional spectrum is complex and often there are emotional conflicts within the sentences. The display of appropriate images is more difficult and the rule is adopted that in such cases will display the appropriate image assigned to the AWEMI. The possible conflicting emotional content in sentences is resolved by this way. With the AWEMI

calculation, the emotion extraction engine is able to display the mood of every user on the screen. It is helpful for a newcomer in the chat room to judge other users' moods instantly.

Rule 26. Conflicting emotions

When a sentence contains both positive and negative emotions, conflicting emotions occur. When conflicting emotions are found, the engine will check the mood calculated previously by the average mood calculation. If the result shows that the user was recently in a specific mood, that sentence's emotion will be set to that particular category with the lowest intensity, otherwise the conflicting emotions will remain unchanged. For example, the sentence "It is a bad day but I want to leave happily" will be considered to be its corresponding historic emotion category if the user was in a happy mood or a sad mood before. If the user's mood is not clear, the sentence's emotion will be happiness plus sadness.

4.4.3.3 The working procedure of the parser

The working procedure of the parser is shown in figure 19. The first step of the parser is to perform the rewrite rule analysis. The aim is to identify the possible combinable phrases and sentence structures. The second step of the parser is to apply the nineteen rules (rule 6 to rule 26), which have been discussed above. According to the function of the detected emotional word, the rewrite rule analysis is slightly different. The following examples illustrate the difference.

- **Emotional adjectives**

If an emotional adjective is found, the parser will search for the participant and tense of the emotional adjective. If a word with opposite meaning is found during this examination, the emotion category will be inverted.

- **Emotional nouns**

If the emotion word is a noun or belongs to a noun phrase, the parser will determine the participant information directly from the word itself or the noun phrase. To obtain tense information, the parser will analyse the output of the rewrite rule component. If a word with opposite meaning is found during this analysis, the emotion category will be inverted.

- **Emotional verbs**

If the emotion word is a verb or belongs to a verb phrase, the parser will determine the tense information directly from the word itself or the verb phrase. To obtain the participant information, the parser will also analyse the output of the rewrite rule component. If a word with opposite meaning is found during the analysis, the emotion category will be reversed.

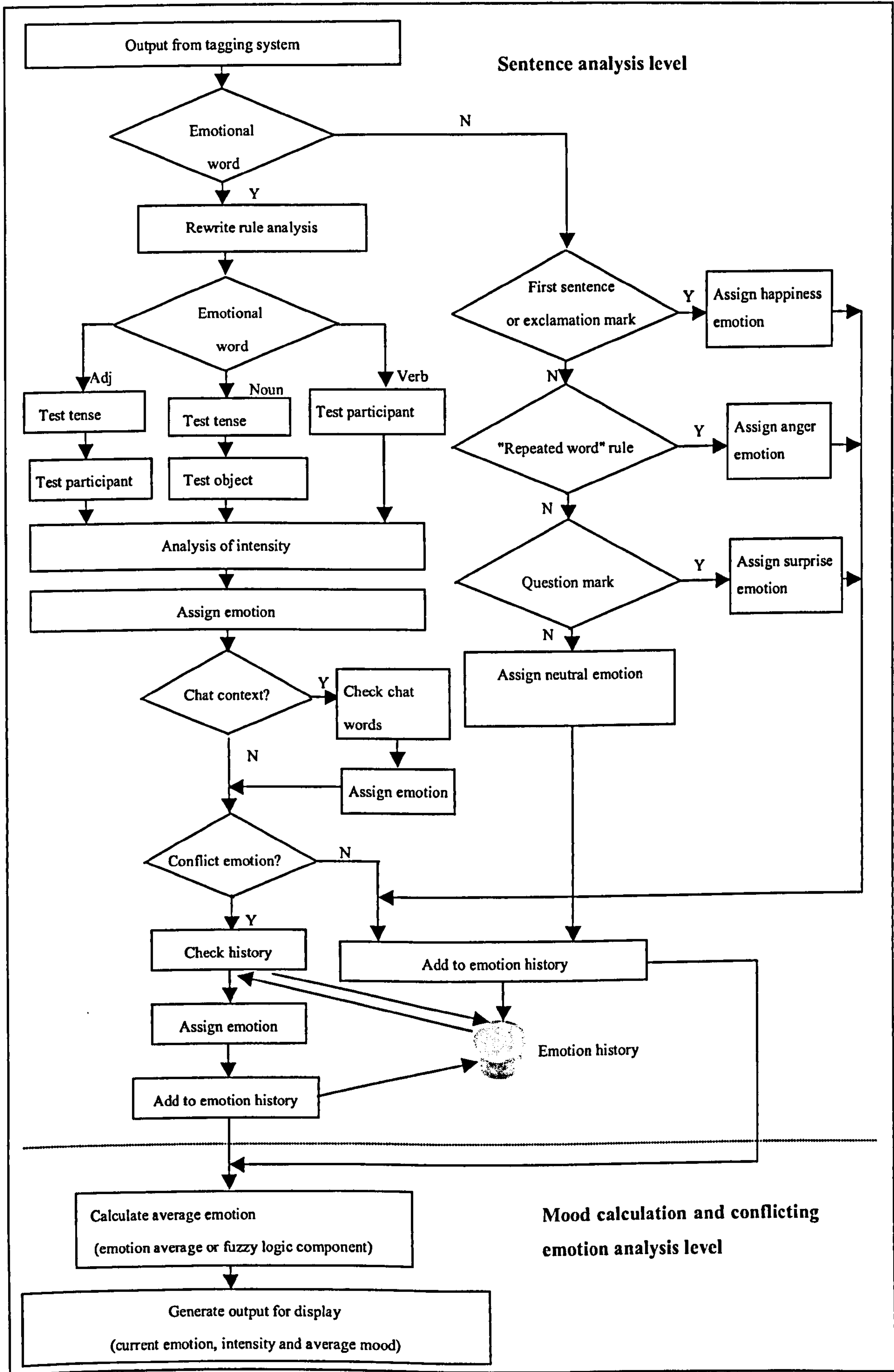


Figure 19: The working procedure of the parser

The output from the parser will be sent to the interface layer. The output contains two components. One is the current emotion of the user, which includes the emotion category, the intensity and the tense and its format is [emotional category] [intensity] [tense]. The other component of the parser's output is the average mood of the user. The average mood includes the emotion category and the intensity and the format is [emotional category] [intensity].

4.4.4 The limitations of the emotion extraction engine

Although various rules and algorithms applied in the emotion extraction engine, the engine still can not handle the complexity of English language and the limitations of the emotion extraction engine are summarised in this section. However, the emotion extraction engine still can be applied within the Internet communication context and achieve considerable performance. As the technology develops, some of the present limitations may be solved in the near future.

1. Sentence context limitation

Context strongly influences the feelings in it. In some specific situations, a sentence containing emotional words does not present any emotional feelings. For example, the sentence "Research shows that happy images can cause sad feelings." does not provide any emotional feelings although emotion keywords are present. The emotion extraction engine aim to solve this problem by detecting the subject the emotional feelings referred to. However as the vast variety of subjects, this limitation still exists. The emotion extraction engine is particularly weak at handling textual messages in special contexts such as science, and medical study.

2. Irony limitations

Ironic emotions exist in both the chat and article contexts. The emotion extraction engine does not implement any algorithms to detecting ironic feelings, and will fail in analysing these sentences as well.

3. Inferred emotion limitations

As inferred emotions are complex and dynamic, different viewers may perceive different emotions from the same sentence. Real-world knowledge is applied in the emotion extraction engine to handle inferred emotions. However, there are as countless real-world knowledge and language context, and the emotion extraction engine can only analyse sentences with relatively strong emotional clues, e.g. sentences such as "I bought a new car!". Under some circumstances, emotional feelings in one sentence may depend on the conclusions of previous sentence,

previous environment or even current public opinions. The emotion extraction engine was not designed to handle this kind of situation.

4. The limitation of system updating

The meanings of emotional words can be different according to context and time. One example is the emotional word "thrilled". The word itself used to present a fearful feeling but now this word is equal to happy. To solve this problem, the tag set and the dictionary need to be updated either manually or automatically. The rules for detecting emotions may need to be updated in the mean time.

5. Keyword limitation

To detect emotional feelings in a sentence, the emotion extraction engine has to find emotional keywords. However, some emotional sentences may not have any significant emotional keywords, e.g., the sentence "I spent ten days on it and finally I worked it out" may presents a relief and happy feeling, but no emotional keyword can be detected. The emotion extraction engine can not detect the emotion in the sentence where no emotion keywords are presented.

6. The limitations of the tagged dictionary

The tagged dictionary was created solely for this project. The tagging and rating (intensity decision etc) were agreed by a group of staff (a developer, a linguist and a HCI expert) in order to achieve a relatively fair rate. However, it possibly would be more accurate to classify and rate the dictionary by combining the output of existing corpora such as BNC or brown corpus.

7. The limitation in multi-emotion display

The emotion extraction engine does detect multi-emotions presented in a single sentence. If no distinctive features can be found in mixed expressions (e.g. a mixed "happiness and sadness" expression), the emotion extraction engine will always try to display one single expression image by calculating the AWEMI. However, in some cases, with the AWEMI the emotion extraction engine can not chose a single expression image, then the mixed expressions will be displayed separately (e.g. a mixed "happiness and sadness" expression may be shown as a happiness expression followed by a sadness expression).

4.5 Development summary

The first part of this chapter described the development of the prototype of an emotion extraction engine that was created as part of the investigation of the methods of extracting emotion from textual messages. The features that enabled communication and provided access to information are described. The emotions are detected by analysing textual messages and can be presented to users in both textual and visual (expressive image) methods. The remainder of this chapter will focus on a series of experiments that assessed the performance of the emotion extraction engine.

4.6 The performance assessment of the emotion extraction engine

The aim of the emotion extraction engine is to sense the emotions embedded in sentences. The main context that the emotion extraction engine to be applied in is the Internet communication environment, e.g. a text based chat room, thus the speed of emotion extraction engine and the accuracy of the emotion extraction engine are the most important performance factors. The development of the emotion extraction engine leads to following two questions that need to be answered:

- Is the emotion extracted by the engine correctly reflecting the intended emotion contained in the sentence?
- Is the emotion extraction engine fast enough to be applied in the real-time Internet communication context?

Two experiments were set up to answer these questions. The first experiment – accuracy experiment examined the accuracy of the emotions detected by emotion extraction engine and the next experiment – speed experiment examined whether the engine is fast enough for the real-time Internet communication context.

4.6.1 Method of the accuracy experiment

The main aim of the first experiment was to determine whether the detected emotions correlate with the reader's understanding. To assess the accuracy of the emotion extraction engine, sentences containing different emotional feelings with different intensities were fed into the emotion extraction engine. The extracted emotions by the engine were compared with the emotions identified manually by the experiment participants.

For Internet communications, text messages can be classified into real-time text messages, i.e., chat based text and non-real-time text, i.e., article-based (e.g. authors may publish articles containing emotional feelings on Internet pages). Sentences presented in published articles are more complex both in sentence structure and in length comparing to chat messages. Second, a published article, e.g., journal articles and news paper articles etc, normally present emotions using conventional words such as “happy” or “sad”. However, in chat messages, slang emotional words (e.g. lol) are widely used.

As the characteristics are different, the experiment was divided into two tests. One was the chat environment test, which text collected from Internet chat rooms were fed into the emotion extraction engine; the other was the article analysis test, in which sentences from different published articles were used. For both chat environment test and article analysis test, the measurement was the amount of the correctly extracted emotional feelings. This was assessed by comparing the extracted emotional feelings with manually extracted feelings.

In the chat environment test, chat archives containing sentences input in real-time were collected from the Internet. All sentences interpreted as emotional were highlighted manually. The chat sentences were sent into the emotion extraction engine and the engine's output was logged. In the article analysis test, articles were collected from the Financial Times website and the BBC web site. In a similar manner to the chat environment test, all sentences interpreted as emotional were highlighted manually. Subsequently these paragraphs were fed into the emotion extraction engine to examine how well its automated estimation correlated with the manual extraction. The chosen articles covered a range of topics including the stock market, accidents and crime.

4.6.2 The operation of the accuracy experiment

In the chat environment test, twenty-three pages of chat logs, which includes six hundred and thirty eight sentences, were collected and input into the engine. For the article analysis test, seven articles, which include three hundred and forty eight sentences, were typed into the emotion extraction engine.

The experiment focused on assessing the ratio of the correctly analysed sentences that the emotion extraction engine was designed to analyse. In this case, sentences belonging to the limitations described in section 4.4.4 were excluded. As conflicting emotions were assessed in a separate experiment, sentences containing conflicting emotions were excluded from this

experiment as well. Totally five hundred and ninety five sentences collected from the chat log and two hundred and fifty three sentences from articles were input into the emotion extraction engine.

4.6.3 Results of the accuracy experiment

For the chat environment test, seventy-eight sentences (13% of the total sentences) were not correctly recognised. For the article test, sixty-one sentences (24% of the total sentences) were not correctly recognised. Figures 20 and 21 showed the experiment result.

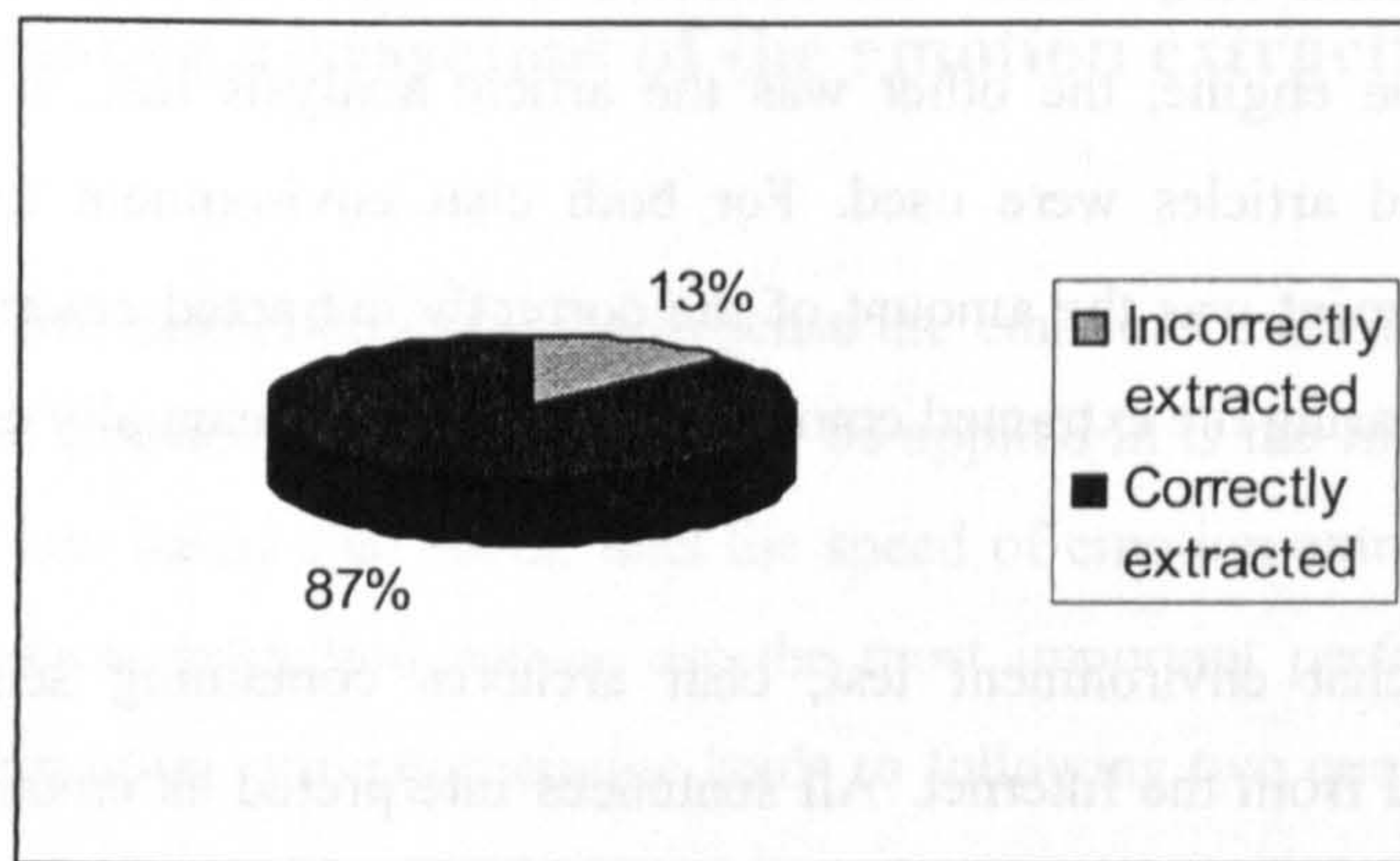


Figure 20: Result of the accuracy test for the chat environment

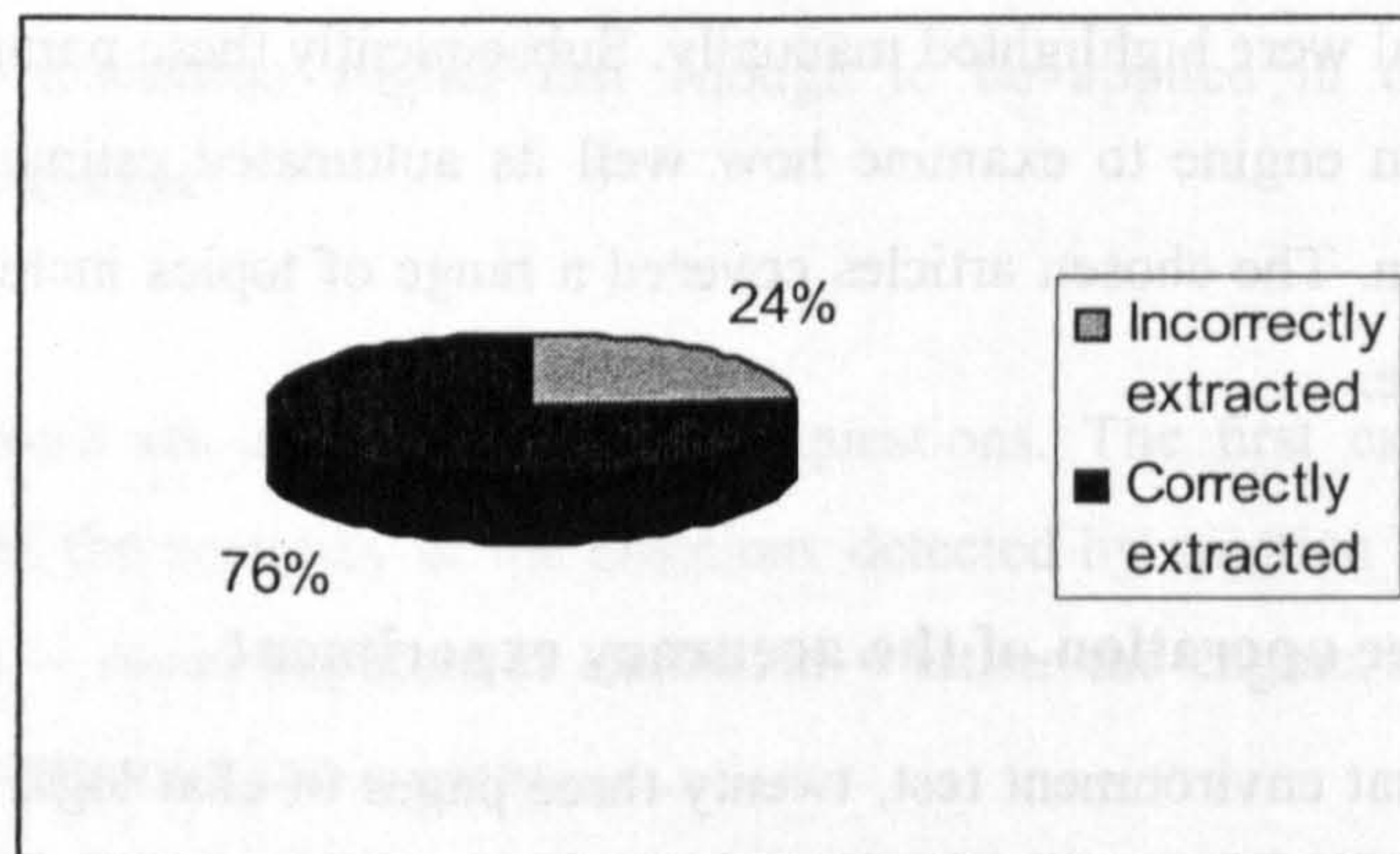


Figure 21: Result of the accuracy test for the article analysis environment

4.6.4 Accuracy experiment result analysis

The results of the tests show that the engine produced better results in the chat environment. The reason is that the sentence structures are much simpler in the chat environment and the individual that the emotion refers to is also limited. However, the results demonstrated that the emotion extraction engine can cope with the majority of emotional sentences in both environments.

The emotion extraction engine achieved more than 75% accuracy in the experiment. This demonstrates that the engine can be used in the real-time environment. However as the sentences that the emotion extraction engine did not designed to handle was excluded, the accuracy of the engine would be lower in possible field tests.

Even with the limitations presented, the experiment can still demonstrate the emotion extraction ability of the emotion extraction engine and the following conclusions can be made. First, the engine may correctly identify emotions in majority sentences when applied in a chat environment. Second, the engine may analyse large numbers of sentences correctly when applied in the article analysis context.

The experiment results illustrate that the emotion extraction engine can be applied in real-time communication context and article analysis context. The next experiment focused on the second question that challenged the performance of the emotion extraction engine, i.e., whether the emotion extraction engine is fast enough to be applied in real-time communication?

4.6.5 Engine speed experiment operation

The experiment was designed to assess the speed performance of the emotion extraction engine. The total time consumed to extract and display visualised emotion (T_{total}) can be divided into three parts, i.e., the time for the engine to analyse the sentences (T_c), the time for network transmission (T_t) and the processing time of the server to send messages out (T_s). Correspondingly $T_{total} = T_c + T_t + T_s$.

The tests were conducted within the 100MB University Intranet. The emotion extraction engine was installed on an Intel P4 computer with 128MB memory and a server to receive and send textual messages and emotion parameters was installed on an Intel P2 computer with 64MB memory as the workload on the server was light. The T_{total} , T_c , and T_s were recorded and the tests were carried out three times to get the average results. The collected results are shown in table 3.

	T_{total} (ms)	T_s (ms)	T_c (ms)
Test 1	62	1	17
Test 2	64	2	18
Test 3	61	1	17
Average	62.3	1.3	17.3

Table 3: Results of the speed experiment

It can be shown that the averaged total consumed time is 62.3 ms ($T_{total}=62.3ms$). In average the engine took about 17.3 ms ($T_c=17.3ms$) to analyse the sentences and select expressive images. The server only needs 1 or 2ms to send messages out. As a result, network transmission can be calculated as $T_t = T_{total} - T_c - T_s$ and the average transmission time was 44ms.

4.6.6 Engine speed experiment results and analysis

The server only consumed 1 or 2 ms, so this consumption can be ignored compared to other components. It is obvious that the network transmission spends more time than the analysis part ($T_c \ll T_t$) and t-test results proved that the mean value of transmission time is significantly higher than the engine's analysis time ($p=0.01$). The test was carried out on a 100MB Intranet and so the transmission time will be much longer if the data is sent over the Internet or using a 56K modem or even a 1M ADSL. The results of the emotion extraction engine speed experiment demonstrate that the engine's analysis is not a bottleneck, the server speed is not a bottleneck and data transmission is the most time consuming component.

4.7 Conflicting emotions experiment

4.7.1 Planning the experiment

The aim of the emotion extraction engine was to sense the emotional feelings embedded in sentences. The accuracy of the engine is the most important performance factor. The mood average method particularly handles the mood and conflicting emotion analysis. The question that needed to be answered in this experiment was: Are the conflicting emotions calculated by the mood average method correctly reflecting the intention of the sentence? To test the accuracy of the mood average methods, sentences containing conflicting emotions with different intensities were fed into the emotion extraction engine. The emotions extracted by the engine were manually compared with the emotions identified by participants.

4.7.2 The design of the conflicting emotion experiment

The conflicting emotion test used three stories that were selected from the BBC web site. Each story contained more than five emotional sentences and a final sentence that contains two conflicting emotions. Participants were asked to read each story. After reading the final sentence of each story, participants were presented with two expressive images. Each expressive image represented one of the conflicting emotions that were presented in the final sentence. Participants were asked to select the image that best represented the emotion presented in the final sentence or describe what expressive image would be more suitable. Fifty participants from Bournemouth University performed in this experiment.

4.7.2.1 Conflicting emotion story 1

Story one described a tragic situation. A woman was pregnant when her husband died in an accident. She gave birth to the couple's second child two months after his death. The final sentence in this story was "The birth has been a time of great happiness tinged with real sadness". The experiment participants were asked "What emotions do you perceive from the last sentence?". Participants could choose from happiness, sadness or other.

- **The emotion extraction engine's choice**

This story presented several sentences containing negative emotions. When analysing the final sentence that contains conflicting emotions, the emotion chosen by the AWEMI was sadness with low intensity.

- **The participants' choice**

The participant's choice is shown visually in figure 22. It can be seen that 45% of the participants chose happiness and another 45% chose other. The remaining 10% selected sadness.

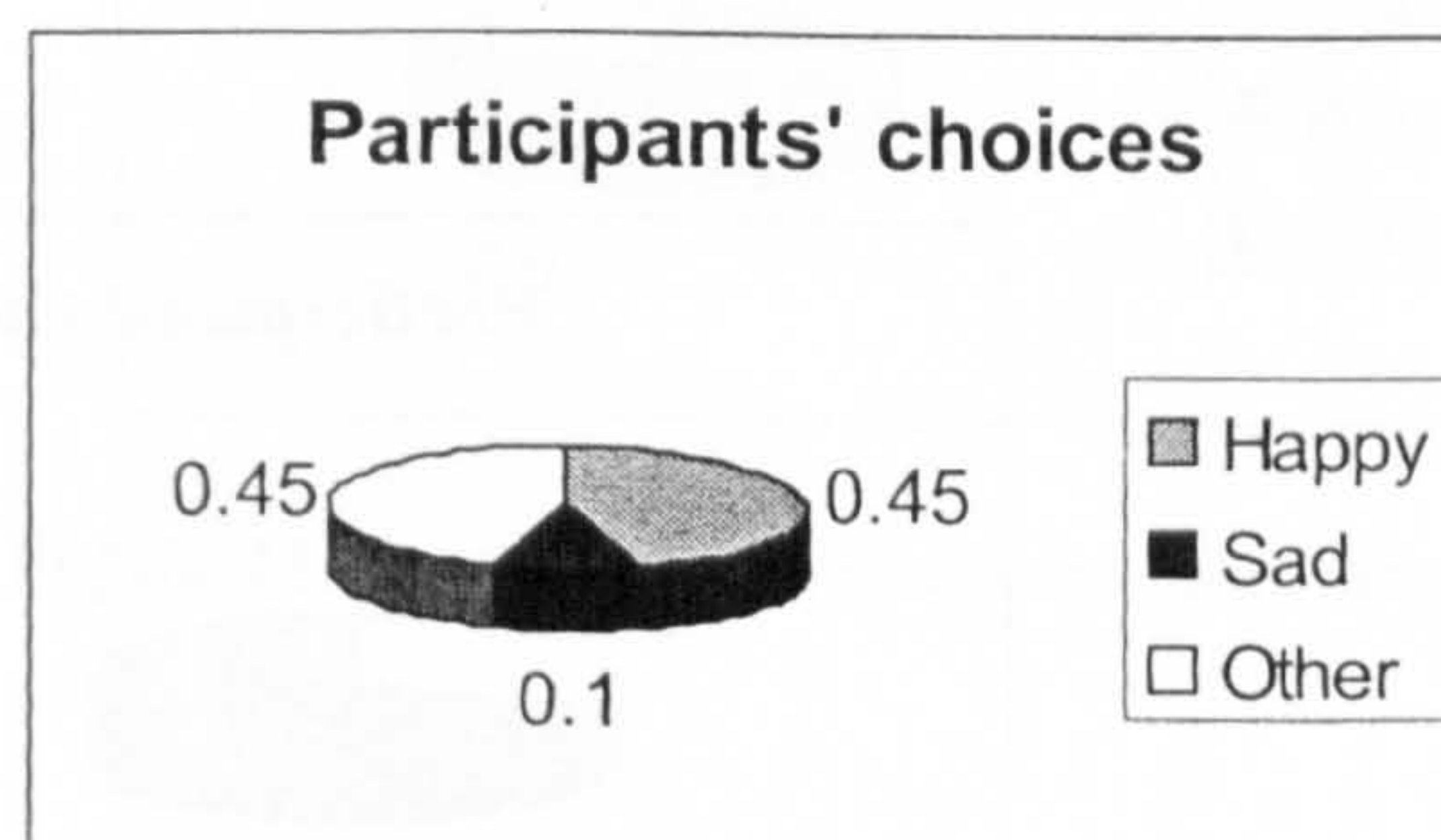


Figure 22: Participants' results for story one

- **Result analysis**

The participants' perceptions demonstrate that the final sentence with conflicting emotions should not be sadness. It should be happiness or at least happiness combined with sadness. The emotion extraction engine chose sadness as more sad emotions were identified in previous sentences. However, more participants believed that the joy of having a baby outweighed the grief for the deceased husband and the baby is the life extension of the husband.

"Having a baby" is such a big event that the happiness it conveyed should be very high. The implications are that some specific issues should be given specific weights to represent the true value. The emotion extraction engine failed to extract the commonly perceived emotion from this sentence.

4.7.2.2 Conflicting emotion story 2

The second story described the life of British soldiers in Iraq. Some soldiers died in the war and a reservist described the disappointment of spending the whole of Christmas in Iraq. The final conflicting sentence was "While the political row over the war rages on, any soldier here can tell you that he just wants to make it home safely. "

- **The emotion extraction engine's choice**

This story presented several sentences with disappointment and sadness emotions. When analysing the final sentence with conflicting emotions, the final emotion decided by engine was anger with lowest intensity.

- **Participants' choice**

The participant's choice is shown visually in figure 23. It can be seen that 56% of the participants chose anger and another 34% chose other. The rest 20% selected happiness.

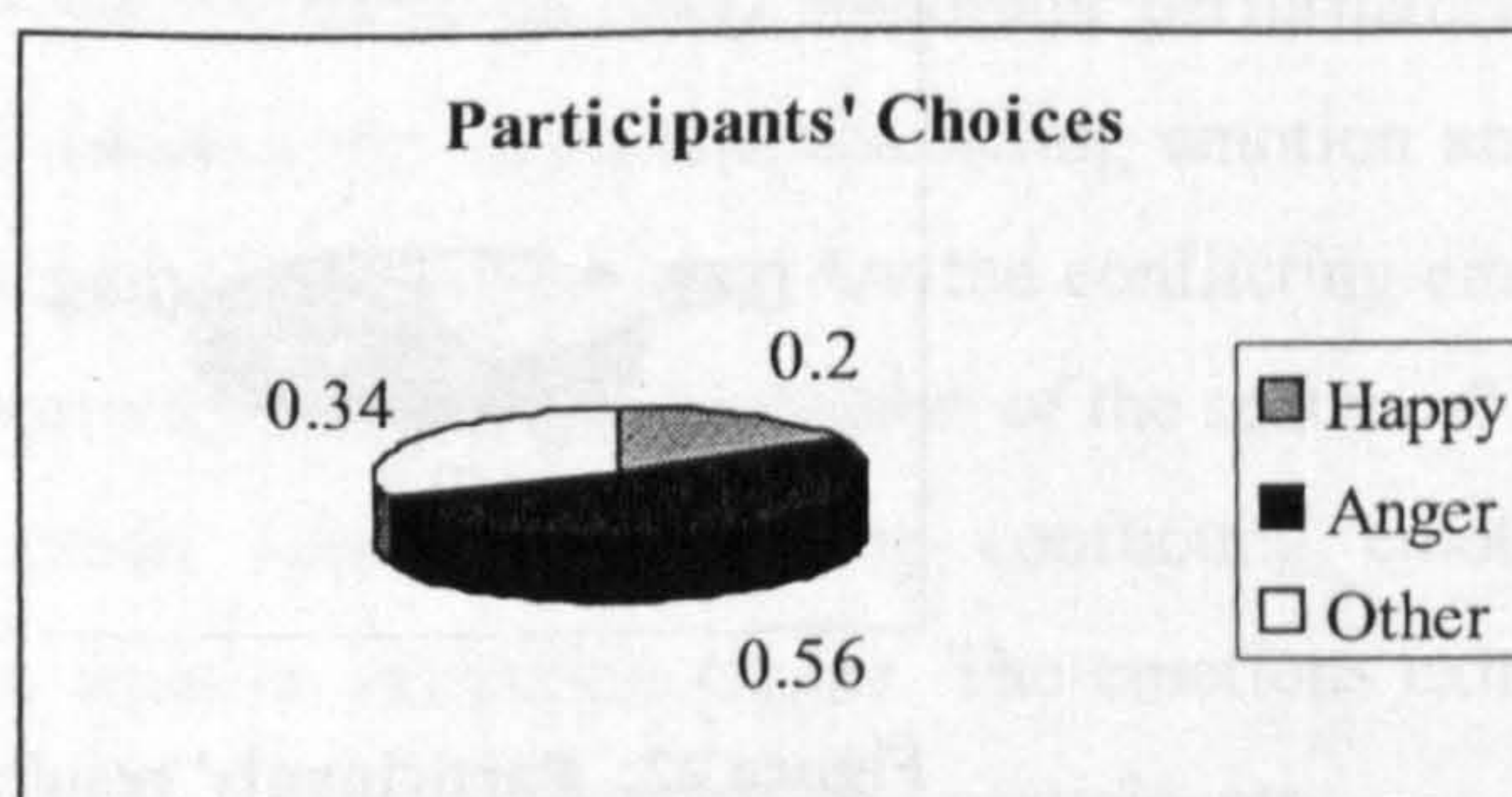


Figure 23: Participants' results for story two

- **Result analysis**

Participants' perceptions demonstrated that an anger feeling was perceived by more participants. The emotion extraction engine chose anger as the emotion for the last sentence. The emotion extraction engine succeeded in identifying the commonly perceived emotion in this story.

4.7.2.3 Conflicting emotion story 3

The third story described a misdiagnosis. A consultant at a hospital told a man in February that he had inoperable pancreatic cancer after complaining of severe abdominal pain. The man resigned from his job the following day and spent £2,500 on a two-week cruise to South America for his last family holiday. When he returned in April, he was told by the hospital that further tests had revealed he did not have pancreatic cancer. The final conflicting sentence was "Initially there was some measure of relief - the anger came later".

- **The emotion extraction engine's choice**

The final sentence presented happiness and anger emotions. The story presented several sentences related to diseases and death which directly link to the sadness emotion. As sadness is a negative feeling, the emotion extraction engine chose the negative feeling - anger to present the emotion of final sentence.

- **Participants' choice**

The participants' choice is shown in figure 24. It can be seen that half of the participants chose anger, 30% chose happy and the rest participants selected "other".

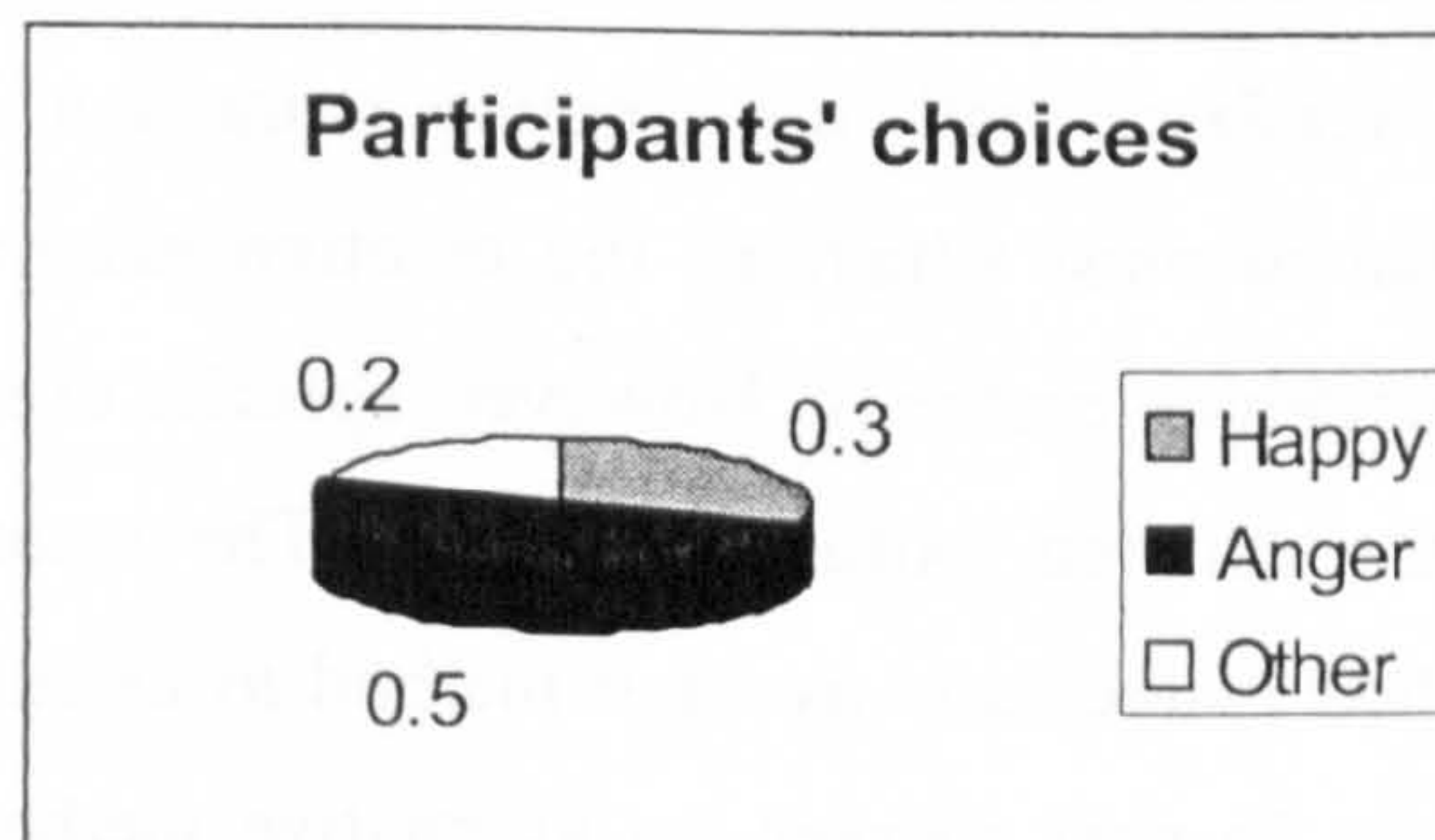


Figure 24: Participants' results for story three

- **Result analysis**

Participants' perceptions demonstrated that the perceived feeling of the final sentence was anger by many participants. The emotion extraction engine chose anger as more negative feelings

were sensed in previous sentences. The emotion extraction succeeded in identifying the correct emotion from this sentence in this story.

4.7.3 Conflict emotion experiment results and analysis

The experiment results show that in two of the three stories, the emotion extraction engine can extract and identify the mood information embedded in textual messages correctly. However, some specific events, e.g. event like “having a baby”, have much stronger impacts on the experiment participants and these events will strongly influence the perceived feelings. This finding shows that sentences describing some specific events should be given greater weight in the emotion analysis. However, as this may depend on the context, the personalities and the mood of the audience, it will require further work to understand and identify relevant events and situations.

4.8 Chapter conclusion

This chapter first described the development of a prototype emotion extraction engine. The emotion extraction engine was designed for real-time Internet communication; however, the engine was tested under both the Internet chat environment and article analysis contexts. To analyse emotions contained in textual messages, a tagged dictionary was created and analysis rules were defined to guide the engine's working flow. The first set of experiments discussed in this chapter concentrated on assessing the two performance issues of the emotion extraction engine: the accuracy and the speed. The experiment results demonstrated that the emotion extraction engine can achieve acceptable accuracy rate and is fast enough to be applied in the real-time communication context. The second set of experiment used three stories and compared the readers' feedback with the emotion extraction engine's output. The experiment results illustrated that in most situations the emotion extraction engine can extract and identify the conflicting emotions correctly. However, specific events (e.g. having a baby) were recognised falsely by the emotion extraction engine. The emotion extraction engine discussed in this chapter applied mathematic average method to calculate the emotion confliction and conduct mood analysis. In next chapter, an alternative method – the fuzzy logic approach to analyse conflicting emotions will be presented.

Chapter 5

Fuzzy emotion analysis functions

The last chapter presented the development of the emotion extraction engine and a set of experiments to assess the engine's performance. To handle mood calculation and conflicting emotion analysis, the mood average method and the fuzzy logic approach were developed. The emotion extraction engine can use either the mood average method or the fuzzy logic analysis functions. In the last chapter, the mood average method was presented and the fuzzy logic analysis functions are discussed here.

The fuzzy functions include four components: conflicting emotion detection component, mood selection component, the emotion storage component and emotion filter component. The working flow of the fuzzy emotion analysis functions are shown in figure 25. In this chapter, the rules to detect conflicting emotions are discussed at the beginning and then the mood selection component will be examined in detail. The 'current emotion states' generated by the mood selection component is stored in the emotion storage component. As the emotion filter only works when enough emotion states are stored in the emotion storage component, the emotion filter component will be discussed after the emotion storage component. To illustrate the workflow clearly, examples will be presented in each component.

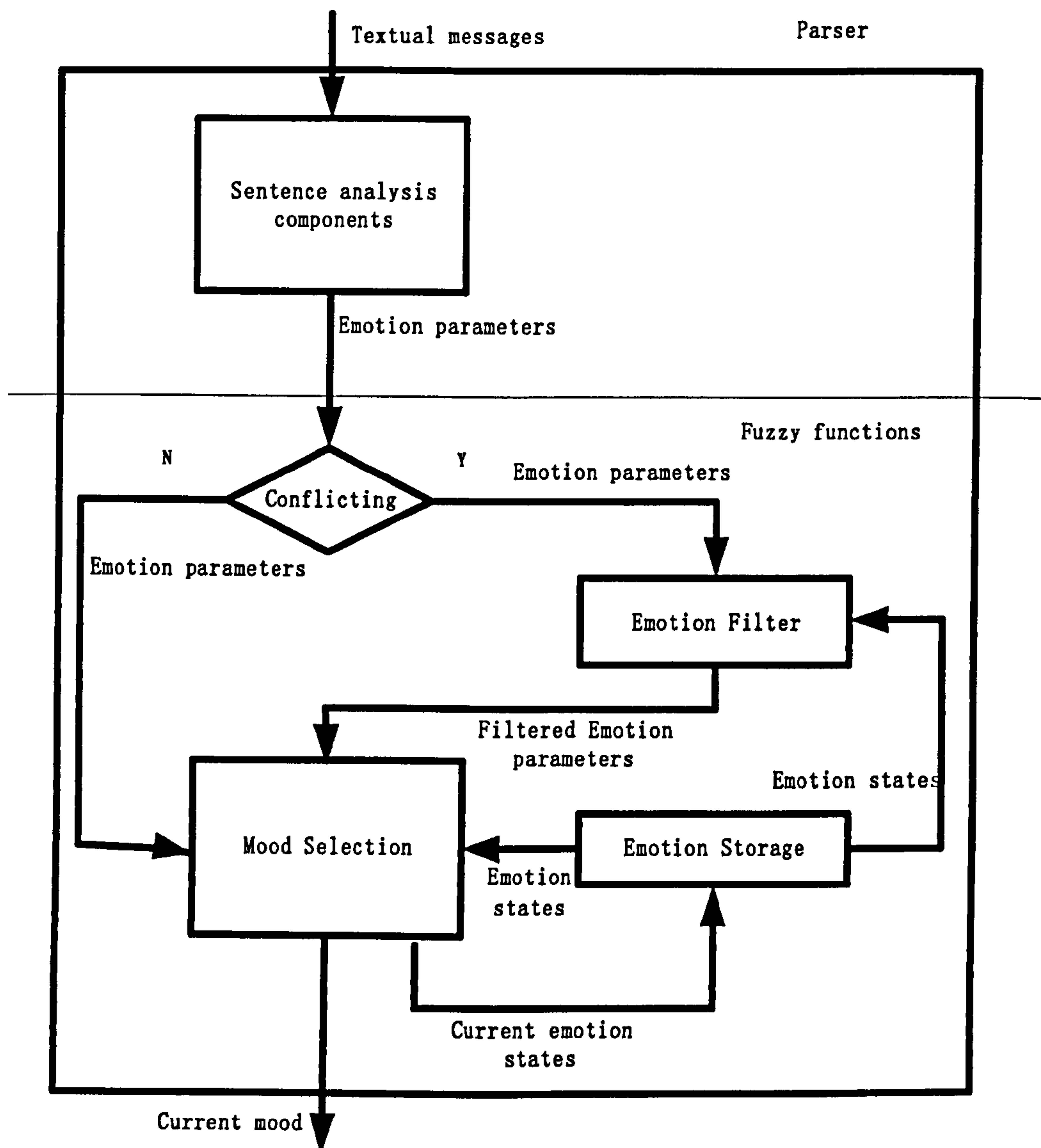


Figure 25: The workflow of the fuzzy emotion analysis functions

5.1 Conflicting emotion detection component

As mixed emotions are a common phenomenon in daily life, it is not unusual for a user to type in a sentence containing mixed emotions in online chat environment, e.g. “I am happy that I got promotion but it is sad that my salary is cut”. Accordingly, it is necessary to have the conflict emotion detection component in the emotion extraction engine.

When a sentence contains conflicting emotions, judging which one can represent the overall emotional feeling is not only based on the current sentence but also on the mood of the conversation. For example, in the emotion extraction engine, the mood “happy” indicates that the previous messages an individual typed in contain overwhelmingly more happy feelings than

others. When the user types a sentence containing both happiness and sadness emotions, the perceived current mood of the user may still be happy, instead of being happy and sad. The reason is that the individual was in a predominately happy mood and the present sadness emotion may not be significant enough to convince the audiences that the mood suddenly changed from happy to sad.

Two general categories: positive emotion category and negative emotion category are introduced to handle conflicting emotions. Positive emotions include happiness and surprise while negative emotions are sadness, fear, anger and disgust. A sentence is treated as a conflicting emotion sentence only if the sentence contains both positive and negative emotions. When a sentence with conflicting emotions is found, the emotion filter component will be invoked. Otherwise the emotion parameters are passed to the mood selection component for further operation.

5.2 Mood selection component

The inputs of the mood selection component are the emotion parameters from the sentence analysis components and the history emotions stored in the emotion storage component. The aim of the mood selection component is to determine the current mood. To achieve this goal, the first step of the mood selection component is to convert the emotion parameters into the current emotions by filtering the tense information, e.g. the emotion parameter [happiness][middle intensity][present tense] is converted to the current emotion [happiness][middle intensity]. The current emotion will then be sent to the storage component as well. Three steps, i.e. emotion array calculation, fuzzy rule analysis and decision making are carried out to choose the mood. An example of the input emotion parameters sent to the mood selection component is shown in figure 26, in which listed the changes of the emotion array over 5 time periods, i.e., from current time to time -5. The range of the emotion intensity is defined from 0 to 3, and 3 is the highest value.

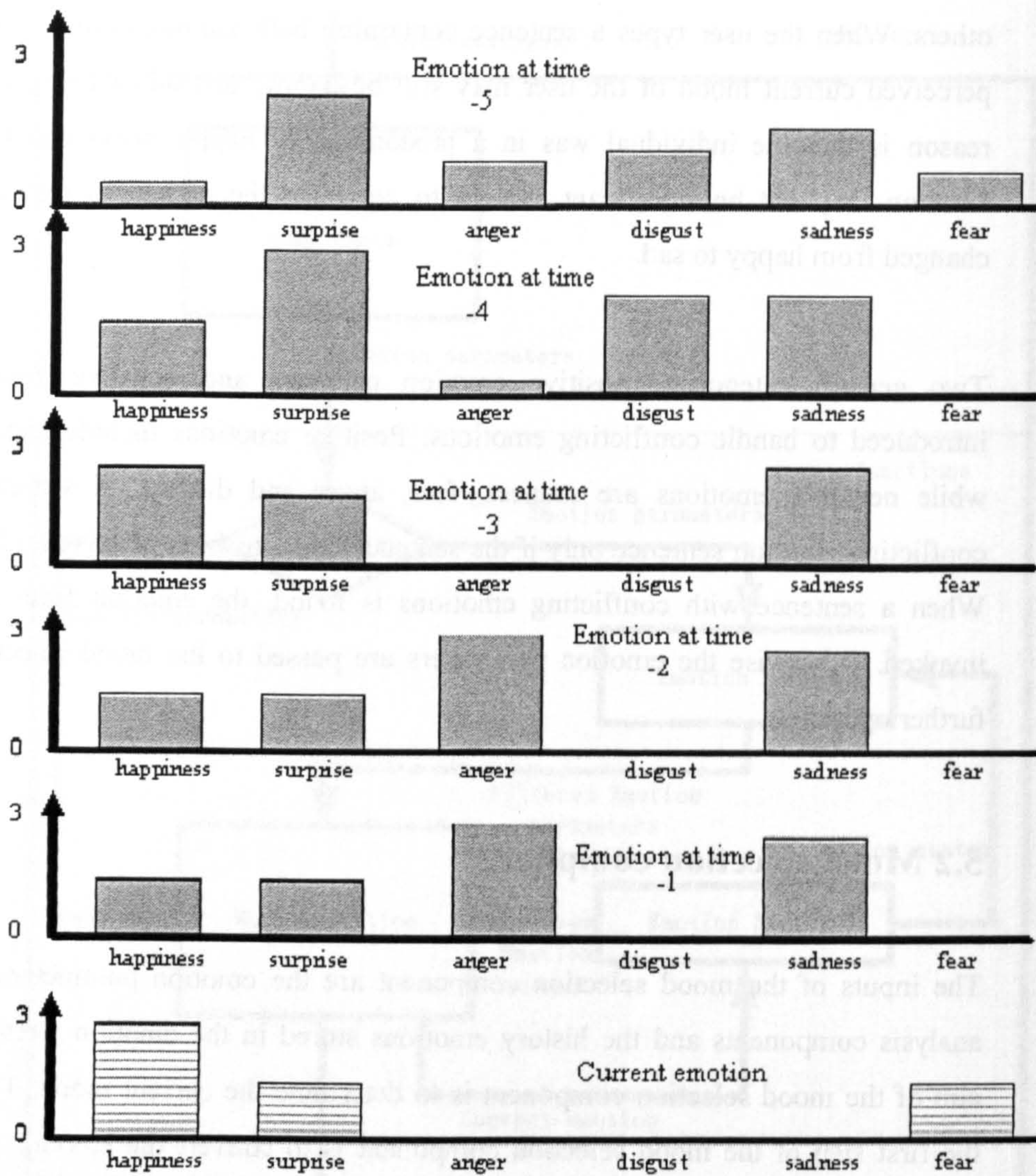


Figure 26: An example of mood selection input

Step 1. Emotion array E[x] calculation

An array E is assigned to store the accumulative intensity values of the six emotion categories. The array elements 0 to 5 in turn represent the accumulative emotion intensity of happiness, surprise, anger, disgust, sadness and fear.

The value of each element in array E is calculated by adding the five previous intensities of a specific emotion category with the current intensity of that emotion, which is shown in the following formula.

$$E[x] = \sum_{t=-n}^0 \alpha_t I_t(x) \text{ where } x = 0,1,2,3,4,5 \tag{5.1}$$

The values of each element in array E depend on the relative intensity over the last n time periods. “n” is chosen to be 5 as it is assumed that in a chat environment, users only remember the most recent dialogs. $I_t[x]$ is the intensity of emotion category x at discrete time t and the value of $I_t[x]$ varies from 0 to 3, which represents the lowest intensity to highest intensity.

- **parameter α**

Instead of adding up the un-weighted previous intensities, the intensities are weighted according to the time. Velasquez [1997] declared that emotions do not disappear once their cause has disappeared, but decay through time. In the FLAME project, El-Nasr and colleagues [El-Nasr et al. 2000] followed Velasquez’s view and chose to decay positive emotions in a faster rate. However, there is not enough empirical evidence to support the decay rate. In the fuzzy logic components, the positive and negative emotions decay at the same rate and the influence period is chosen to be the previous five sentences. Figure 27 illustrates the assumption.

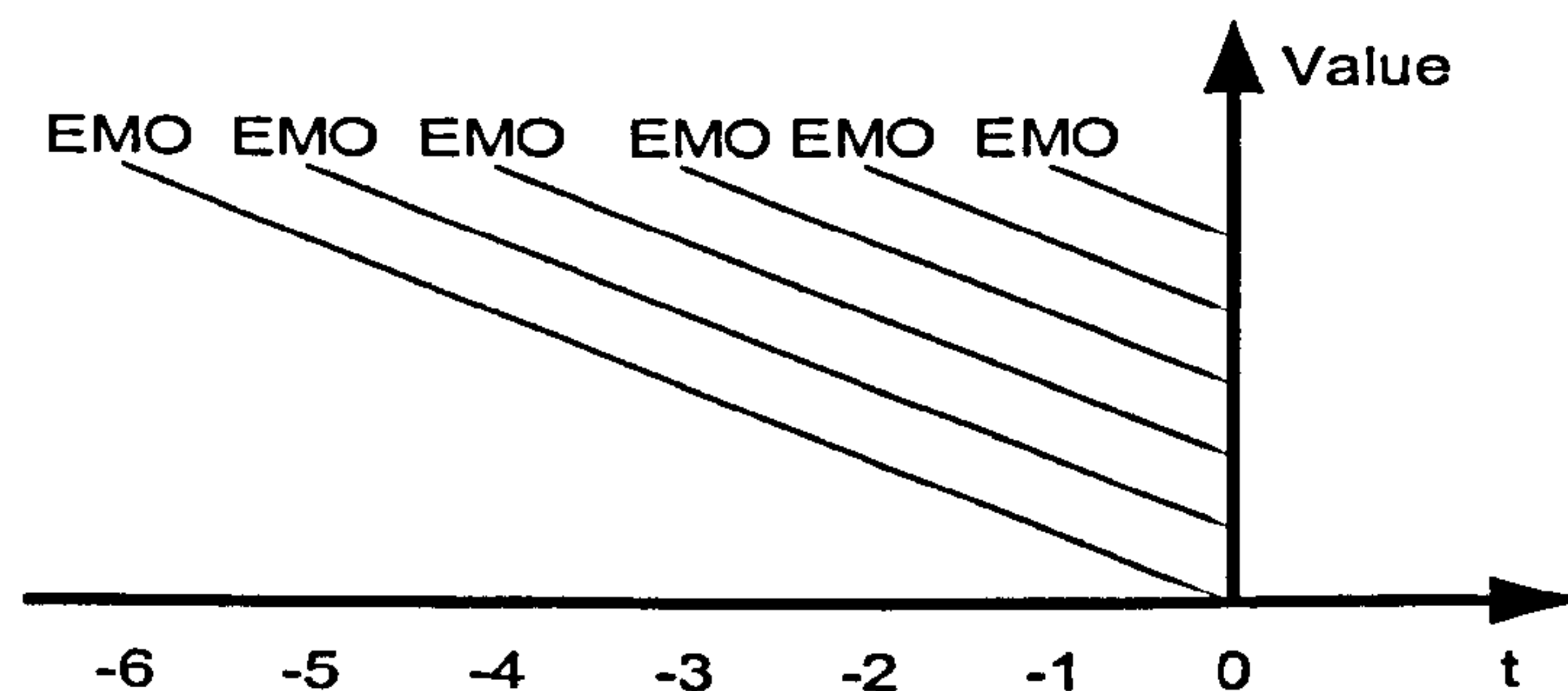


Figure 27: The decay of emotion over time

In the above figure, "value" represents the intensity of perceived influence of emotion and t represents time. EMO represents emotion occurred at the discrete time point (e.g. the time when a chat message is typed in). The assumption of this figure is that time 0 represents the current time. In this figure, the values of the EMO at different time are the same, which means that a user typed six sentences containing the same emotions and the same emotion intensities.

However, at time 0, the perceived influence of EMO occurred at time -6 is 0, which means the influence of emotion at time -6 has disappeared. EMO occurred at time -1 is the least decayed and has the strongest influence to current time (time 0). In the mood selection component, the emotion decay is represented by the weight parameter α and is calculated using formula 5.2.

$$\alpha_t = \alpha_{t+1} - 0.1 \text{ where } t \in -5, -4, -3, -2$$

$$\alpha_{-1} = 0.5$$

$$\alpha_0 = 1$$

(5.2)

• **Emotion array E[x] calculation example**

The following example illustrates the procedure to calculate the values of array E. First, the intensities I[x] is calculated and then weighted by parameter α . Second, the emotion array E is calculated by summing up the corresponding weighted I by time t, e.g. E[0] (0 represents emotion happiness) equals to the sum of weighted intensities I[0] from time -5 to -1. Table 4 to table 6 present an example of the emotion array E calculation. In table 4, the raw intensities for each emotion category from discrete time 0 to -5 are presented. Table 5 lists the weighed values of each element of array E. In table 6, the values of array E[x] is calculated by adding the weighted intensities from time 0 to time -5.

Time	I _t [0] Happiness	I _t [1] Surprise	I _t [2] Anger	I _t [3] Disgust	I _t [4] Sadness	I _t [5] Fear
t=-5	I ₅ [0]=1	I ₅ [1]=2	I ₅ [2]=0	I ₅ [3]=0	I ₅ [4]=0	I ₅ [5]=0
t=-4	I ₄ [0]=0.5	I ₄ [1]=1	I ₄ [2]=0	I ₄ [3]=0	I ₄ [4]=0	I ₄ [5]=1
t=-3	I ₃ [0]=1	I ₃ [1]=0.66	I ₃ [2]=0	I ₃ [3]=0	I ₃ [4]=0	I ₃ [5]=0
t=-2	I ₂ [0]=0	I ₂ [1]=0.25	I ₂ [2]=0.25	I ₂ [3]=0	I ₂ [4]=0	I ₂ [5]=0
t=-1	I ₁ [0]=0.8	I ₁ [1]=0	I ₁ [2]=0.2	I ₁ [3]=0	I ₁ [4]=0	I ₁ [5]=0.4
t=0	I ₀ [0]=0	I ₀ [1]=0	I ₀ [2]=0	I ₀ [3]=0.2	I ₀ [4]=0.2	I ₀ [5]=0

Table 4: The raw intensity values

Time	$\alpha_t I_t$ [0] Happiness	$\alpha_t I_t$ [1] Surprise	$\alpha_t I_t$ [2] Anger	$\alpha_t I_t$ [3] Disgust	$\alpha_t I_t$ [4] Sadness	$\alpha_t I_t$ [5] Fear
t=-5	0.1	0.2	0	0	0	0
t=-4	0.1	0.2	0	0	0	0
t=-3	0.3	0.2	0	0	0	0
t=-2	0	0.1	0.1	0	0.1	0
t=-1	0.4	0	0.1	0	0	0.2
t=0	0	0	0	0.2	0.2	0

Table 5: The weighted intensities

	E[0] Happiness	E[1] Surprise	E[2] Anger	E[3] Disgust	E[4] Sadness	E[5] Fear
Value	0.9	0.7	0.2	0.2	0.3	0.2

Table 6: The values of array E

Step 2. Fuzzy rule analysis

The second step of the mood selection component is to create fuzzy membership functions and carry out fuzzy rules analysis. To analyse the mood of a user, it is necessary to calculate the intensities of the more general categories: positive and negative emotions. Positive emotions (emotion happiness and surprise) are mapped into the elements 0 and 1 of the emotion array. Negative emotions are mapped into the rest of the elements of the array. The intensities of positive and negative emotions are calculated using formula 5.3 and 5.4.

$$Pos = \sum_{x=0}^1 E[x] \tag{5.3}$$

$$Neg = \sum_{x=2}^5 E[x] \tag{5.4}$$

Pos is the intensity of positive emotions and *Neg* is the intensity of negative emotions. For the above example, *Pos* is calculated as 1.6 and *Neg* equals to 0.9. One fuzzy membership functions "Percent" is defined to calculate the percentage of each element in the emotion array E. This function is defined in formula 5.5.

$$Percent(x) = \frac{E[x]}{\sum_{i=0}^5 E[i]} \tag{5.5}$$

- **Membership function example**

The values of membership function "Percent" for each element of array E is shown in table 7.

	Percent (E[0])	Percent (E[1])	Percent (E[2])	Percent E[3])	Percent (E[4])	Percent (E[5])
Value	0.36	0.28	0.16	0	0.04	0.16

Table 7: The "Percent" values of array E in the example

- **Fuzzy rules**

Rule 1

The corresponding emotion category of the element in emotion array E with the largest value is chosen to be the current mood and the output value is the value of its membership function “Percent”.

The elements in emotion array E (elements are from 0 to 5) represents the different emotion categories and their values are the accumulative weighted intensities of corresponding emotion categories. This rule actually chooses the emotion with the largest accumulative weighted intensity to be the current mood. The value of the output – “the value of truth for rule 1” is the percentage of the chosen emotion’s accumulative weighted intensity in all the emotions’ accumulative weighted intensity. For example, if E[1] is chosen to be the current mood and the output value is 0.4, it will mean that emotion “surprise” has the largest accumulative intensity and the percentage of emotion “surprise” in all the presented emotions is 40%. The value of truth for this rule is 0.4.

Rule 2

The positive emotion is chosen to be the current mood and the output value – “the degree of truth” is the value of its membership function “Percent”.

Rule 2 selects the positive emotion as the current mood no matter how the emotions were actually presented. If the positive emotions were presented overwhelmingly, the degree of truth for this rule will be high; otherwise the truth for this rule will be low. This rule is created to handle situations such as emotions happiness and surprise were both presented overwhelmingly.

Rule 3

The negative emotion is chosen to be the current mood and the output value – “the degree of truth” is the value of its membership function “Percent”.

Rule 3 is created to handle situations when more than one emotion in the negative emotion category were presented overwhelmingly. Rule 3 selects the negative emotion as the current mood no matter how the emotions were actually presented. Similar to rule 2, if the negative emotions were presented overwhelmingly, the degree of truth for this rule will be high, otherwise the truth will be low.

• **Fuzzy rule calculation example**

In the latest example, x represents the emotion category and the variable $E[x]$ is the weighted accumulative intensity of emotion category x . For rule 1, $E[0]$ has the largest value and the value of membership function "Percent" is 0.36. As a result, emotion happiness is chosen as the current mood and the output of rule one – the degree of truth of rule 1 is 0.36. Rule No.2 selects positive emotion as the current mood, the positive emotion category's value - Pos is 1.6 and negative emotion category's value - Neg is 0.9. The value of the membership function "Percent" of positive emotion is 0.64. As a result, the output of rule two is 0.64. Rule No.3 chooses negative emotion as the current mood and the value of the membership function "Percent" is 0.36. Table 8 summarises the output values of rule 1 to rule 3.

	Rule 1	Rule 2	Rule 3
Output (the degree of truth)	0.36	0.64	0.36

Table 8: Output of rule 1 to 3

Step 3. Decision making process

The decision making process determines which emotion the mood selection component will choose as the current mood. The decision making process assumes that when a dialog starts, the mood of new user is neutral. The centre of gravity (cog) point is an important measurement factor in the decision making process. In this research work, the centre of gravity (cog) point is calculated as the average of the rules' outputs, which is shown in formula 5.6. Based on the output of rules and cog value, the following four possible decisions can be made.

$$cog = \frac{\sum_{x=1}^3 rule(x)}{3} \tag{5.6}$$

Decision 1

If only rule one's output is over the cog point, then the decision follows that rule's description.

When only rule one's output is over the cog point, there is a situation in which a single emotion was presented overwhelmingly and other emotions were mentioned much less. The mood selection component will choose the current mood to be the corresponding emotion category described in rule one. The intensity of current mood is the same as the weighted accumulative emotion intensity (AWEI).

Decision 2

If only the output rule 2 or rule 3's is over the cog point, then the decision follows rule 2 or 3's description.

When only rule 2 or rule 3's output is over the cog point, it shows that no single emotion was presented overwhelmingly most frequently; however, the general positive or negative emotions were mentioned overwhelmingly. For example, the emotion happiness and surprise were both presented frequently, but the accumulative intensity of happiness and surprise were very close. The mood selection component will choose the general category as the current mood. The mood selection component aims to choose one emotion to present current mood. The emotion with highest intensity in the general emotion category will be chosen as the current mood. It is arguable that the current mood should combine the emotions in the positive or negative categories, however it is extremely difficult to generate acceptable facial expression images to represent the combined feelings. As a result, the mood selection component will just choose one emotion as the current mood.

Decision 3

If the output of rule 1 and either of rule 2 or rule 3's output is over the cog value and rule 1 agrees with that rule, then the decision follows rule 1.

When the output of rule 1 is over the cog point, it means that a specific emotion was presented overwhelmingly. When the output of either rule 2 or rule 3 is over the cog point, it means that either positive or negative emotions overwhelmed the dialog. "Rule 1 agrees that rule" means that the specific overwhelming emotion belongs to the overwhelming general emotion category. It demonstrates that not only the general emotion category X was presented overwhelmingly, but also that the specific emotion X contributes the greatest amount to make the general emotion category overwhelming. In this case, the specific emotion in rule 1 contributes most to the overall feelings and the category of mood of the user is chosen to be the same as the specific emotion.

Decision 4

If the output of rule 1 and either of rule 2 or rule 3's output is over the cog value, but rule 1 does not agree with that rule, then the decision follows the previous mood.

Rule 1 demonstrates that a specific emotion occurred much more than any other emotions, however the accumulative intensity of the opposite emotions can still be high enough to make the opposite general emotion value higher than the cog point.

This illustrates the situation in which specific emotions was presented at an extremely high intensity and opposite emotions were presented in a low intensity and continuous manner. In this situation, the mood of a user was sophisticated and emotion confusion was presented. It is reasonable to assume that mood is not easily changeable, therefore the category of the current mood is the same as the previous mood. However, the intensity of the mood is decreased to be lowest to represent the complex feelings.

Decision 5

If the output of none of the rules are over the cog point, then the decision follows the previous mood.

As no emotions occur significantly more than others, the current mood is kept the same as the previous mood and the intensity is decreased to represent the decay effect. If the previous mood was already in the lowest intensity, then the current mood is set to be neutral.

- **The decision making example**

Figure 28 shows the cog point of the example demonstrated in this section. It can be seen that only the output of rule 2 is over the COG value and so decision two is taken. The decision two - current mood is the highest emotion in the positive emotion category, which is "happiness" in this example. The result of this example is that the detected current mood is happy and its intensity is 0.8, although the current emotion extracted by the emotion engine (shown in figure 29) only contains two negative emotions "sadness" and "disgust".

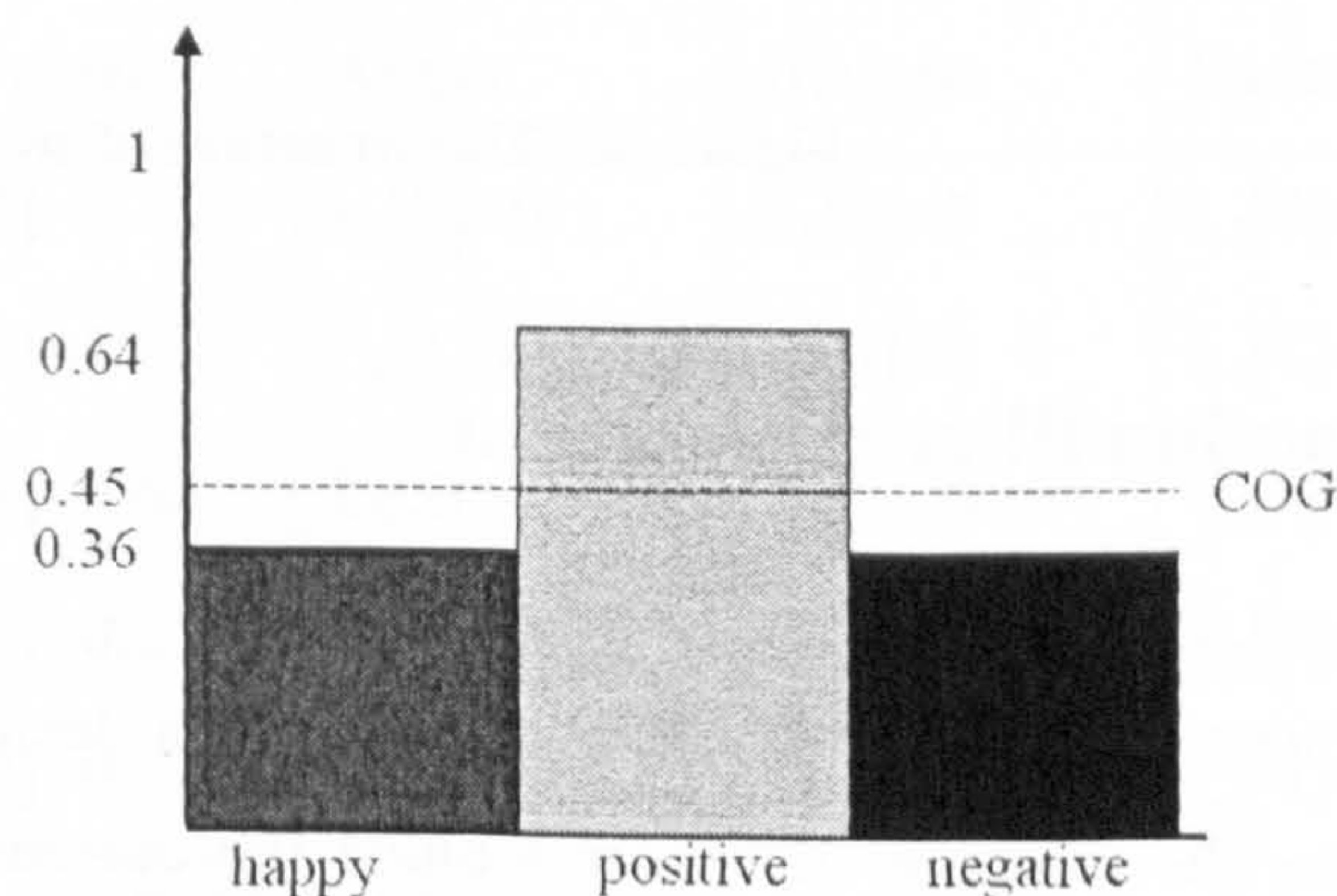


Figure 28: The COG example

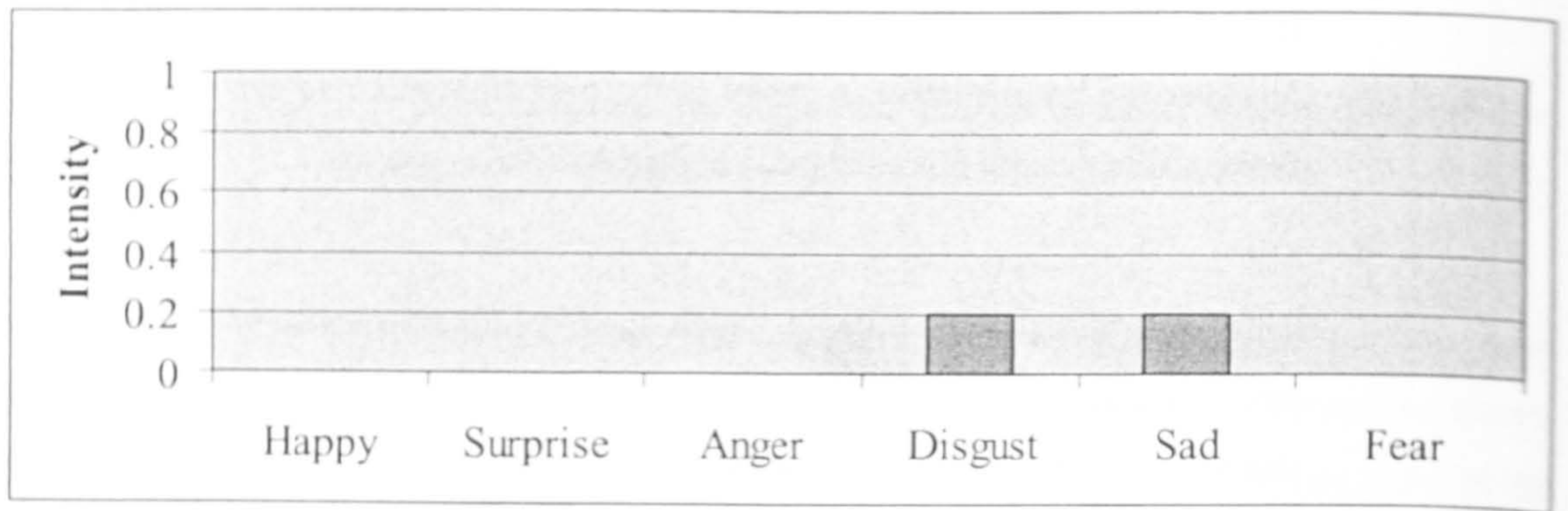


Figure 29: Emotions extracted from the example sentence

5.3 Emotion storage component

The inputs of the emotion storage component are the extracted emotions at different discrete times. The storage component is implemented as a first in first remove (FIFR) stack with a length of five. The structure is shown in figure 30.

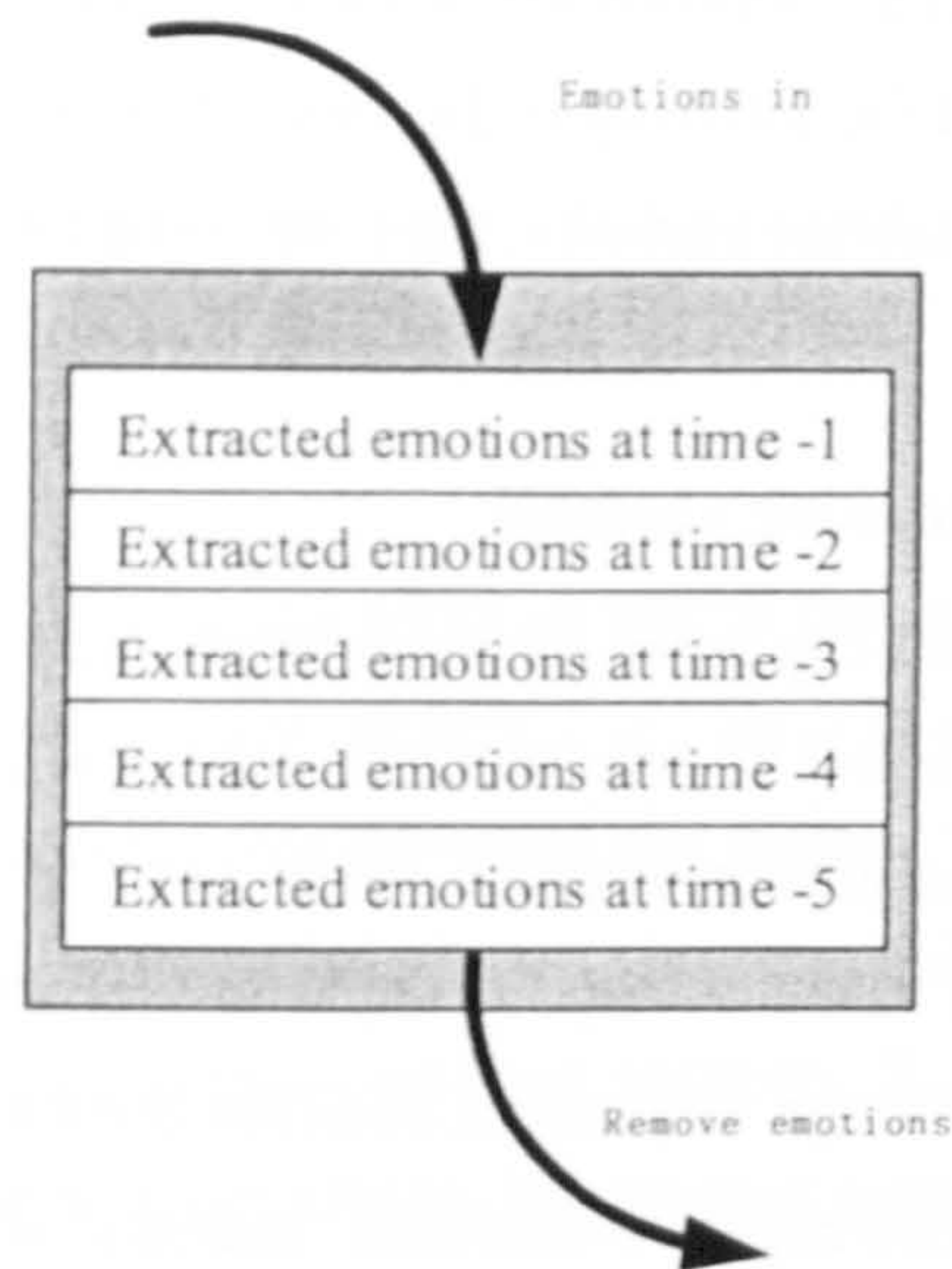


Figure 30: The structure of storage component

5.4 Emotion filter component

The purpose of the emotion filter component is designed to analyse the detected conflicting emotions. The inputs of the filter include the current conflicting emotions and the previous emotions stored in the emotion storage component. Three steps, i.e., FE array calculation, fuzzy rule analysis and decision making process, are carried out to analyse the conflicting emotions.

Step 1. FE[x] array calculation

Similar to the mood selection component, an array FE (filtered emotion) is assigned to contain the cumulative intensity value of emotions:

$$FE[x] = \sum_{i=-5}^{-1} \alpha_i \times I_i(x) \text{ where } x = 0,1,2,3,4,5 \quad (5.7)$$

The weight parameter α and intensity variable $I[x]$ have the same meanings as in the mood selection component (formula 5.1). The difference is that the emotions at time 0 – the current emotions are excluded from the calculation of the elements in the FE array. Positive and negative emotions are calculated using the same formulas (formula 5.3 and 5.4) as in the mood selection component.

- **Example of calculation array FE**

For example, two emotions, happiness and sadness are presented in the same intensity in one sentence; as a result, the conflicting emotions of happiness and sadness are detected. To analyse the emotion conflict, the emotion filter component will check the emotion histories saved in the emotion storage component.

In this example, the history emotion values used in the mood selection component are re-used. In this example, two conflicting emotions (happiness and sadness) are expressed in the same intensity. The raw values of intensity variable $I[x]$ can be found in table 9 and the weighted intensities are listed in table 10. The only difference from the sample used in the mood selection component is that the intensities of current emotion are excluded. By applying formula 5.7, the array FE is calculated and displayed in table 11.

Time	I _t [0] Happiness	I _t [1] Surprise	I _t [2] Anger	I _t [3] Disgust	I _t [4] Sadness	I _t [5] Fear
t ₋₅	L ₅ [0]=1	L ₅ [1]=2	L ₅ [2]=0	L ₅ [3]=0	L ₅ [4]=0	L ₅ [5]=0
t ₋₄	L ₄ [0]=0.5	L ₄ [1]=1	L ₄ [2]=0	L ₄ [3]=0	L ₄ [4]=0	L ₄ [5]=1
t ₋₃	L ₃ [0]=1	L ₃ [1]=0.66	L ₃ [2]=0	L ₃ [3]=0	L ₃ [4]=0	L ₃ [5]=0
t ₋₂	L ₂ [0]=0	L ₂ [1]=0.25	L ₂ [2]=0.25	L ₂ [3]=0	L ₂ [4]=0	L ₂ [5]=0
t ₋₁	L ₁ [0]=0.8	L ₁ [1]=0	L ₁ [2]=0.2	L ₁ [3]=0	L ₁ [4]=0	L ₁ [5]=0.4

Table 9: The raw intensity values

Time	$\alpha_t \times I_t[0]$ Happiness	$\alpha_t \times I_t[1]$ Surprise	$\alpha_t \times I_t[2]$ Anger	$\alpha_t \times I_t[3]$ Disgust	$\alpha_t \times I_t[4]$ Sadness	$\alpha_t \times I_t[5]$ Fear
τ_5	0.1	0.2	0	0	0	0
τ_4	0.1	0.2	0	0	0	0
τ_3	0.3	0.2	0	0	0	0
τ_2	0	0.1	0.1	0	0.1	0
τ_1	0.4	0	0.1	0	0	0.2

Table 10: The weighted Intensities

FE[0]	FE[1]	FE[2]	FE[3]	FE[4]	FE[5]
0.9	0.7	0.2	0	0.1	0.2

Table 11: The values of array FE in this example

Step 2. Fuzzy rule analysis

The Pos and Neg values are calculated in the same manner as in the mood selection component. The same fuzzy membership function "Percent" in the mood selection component are used in the emotion filter component. The difference is that in the emotion filter, the membership functions "Percent" is applied to the general emotion categories positive and negative emotions, not to individual emotion categories. To analyse conflicting emotions, following two rules are made.

Rule 1

The negative emotion in the conflicting emotions are removed and the value of positive emotion category's membership function "Percent" is chosen to be the degree of truth of rule 1. The intensity of the positive emotion in the conflicting emotions is decreased.

Rule 1 selects the positive emotion as the emotion to represent the feelings of the conflicting emotions no matter how the emotions were actually presented. If the positive emotions were presented overwhelmingly before, the degree of truth for this rule will be high; otherwise the degree of truth for this rule will be low.

Rule 2

The positive emotion in the conflicting emotions are removed and the value of negative emotion category's membership function "Percent" is chosen to be the degree of truth of rule 2. The intensity of the negative emotion in the conflicting emotions is decreased.

In the same manner as for rule 1, if the negative emotions were presented overwhelmingly before, the degree of truth for this rule will be high; otherwise the degree of truth for this rule will be low.

- **Fuzzy rule calculation example**

The above example's *Pos* and *Neg* values are 1.6 and 0.5. The output values of both rules – the degree of truth of membership functions "Percent" for positive and negative emotions are calculated and shown in table 12.

	Percent(<i>Pos</i>)	Percent(<i>Neg</i>)
Value	0.762	0.238

Table 12: The "Percent" membership's values of *Pos* and *Neg* in the example

Step 3. Decision making process

The Centre of Gravity (COG) point is calculated as the average point of the rules' outputs as in formula 5.8.

$$COG = \frac{\sum_{x=1}^2 rule(x)}{2} \tag{5.8}$$

If a sentence containing conflicting emotions is the first sentence a user typed in, the emotion filter will not delete any emotions because no history emotion is stored in the emotion storage component. When more data arrives, e.g. the user start chatting with others, the history information can be created and above fuzzy rules can be applied. Based on the output of fuzzy rules, the following decisions can be made.

Decision 1

If one rule's output is above the COG point, then the decision follows that rule.

When one rule's output is above the COG point, it means that either positive or negative emotions were overwhelmingly presented. The current emotions that conflict with the overwhelming general emotion category are removed and the intensities of the emotions that remain in the sentence are decreased.

Decision 2

If none of the rules' outputs are above the COG point, then the conflicting emotions are kept.

When neither positive nor negative emotions are dominant in a user's input sentences, a complex and contradictory feeling exists. This may mean that the user tends to describe contradictory and ambiguous emotions. As it is assumed that mood is not easily changeable, the conflicting emotions are kept.

- **Decision making example**

For the above example, the COG point is 0.5. As the output of rule 1 is over the COG point, the description of rule 1 is followed and as a consequence the negative emotions are removed.

5.5 Development summary

The first part of this chapter described the development of the fuzzy emotion analysis components for the emotion extraction engine. A set of fuzzy rules are applied to guide the fuzzy component's work. The difference between the mood average method described in chapter 4 and the fuzzy components described in this chapter are fundamental. Conflicting emotions can be retained to represent the mood in the fuzzy logic method (e.g. a user may be in happy and sad mood). In contrast, the mood average method described in chapter 4 always selects one emotion to represent the mood. To assess the performance of the fuzzy emotion analysis components, an experiment was carried out and is presented in following section.

5.6 Conflicting emotion experiment

5.6.1 Experiment planning and methods

The fuzzy emotion analysis components presented an alternative approach in handling mood analysis and conflicting emotion analysis. The question that needed to be answered in this

experiment is: Is there any performance difference between the mood average method presented in last chapter and the fuzzy approach?

To answer this question, an experiment was carried out. In the same manner as in the experiment carried out to detect the performance of the mood average method, three stories were feed into an emotion extraction engine with the fuzzy components. The three stories were chosen to be exactly the same three stories used in the experiment for the mood average method in order to compare the performance difference. For story details and experiment environment details, please refer to section 4.6.

5.6.2 Experiment result analysis

This first story presents several sentences containing negative emotions. When analysing the final sentence that contains conflicting emotions, the fuzzy analysis component determined that the rule 2's output is over the COG point and the chosen emotion is sadness. This second story presented several sentences with disappointment and sadness emotions. When analysing the final sentence with conflicting emotions, only rule 2's output is over the COG point and the emotion is chosen as anger. In story three, sentences with angry feelings are presented and the final sentence presents both happiness and anger emotions. The fuzzy components' output shows that rule 2 is qualified and the emotion was set to be anger.

The experiment results show that in fuzzy logic analysis approach agreed on the results of the mood average method calculated. No difference was found in the comparison. One explanation is that although theoretically different, these two implemented approaches actually achieved the same emotion extraction performance. Another explanation is that the amount of input data (the articles) is not large enough to reveal the possible performance differences between these two methods.

5.7 Chapter conclusions

The fuzzy emotion analysis components are presented in this chapter. The fuzzy components are an alternative method to the mood average method and can be integrated into the parser of emotion extraction engine. To identify emotions, fuzzy logic knowledge was applied and a series of rules were defined to guide the fuzzy component's work. To assess the performance of the fuzzy emotion analysis components, three stories used in the conflict emotion experiment in chapter 4 were input and the feedback of the fuzzy components were compared with the mood

average method's feedback. The experiment results illustrated that the fuzzy method achieved the same performance as the mood average method.

The previous chapter and this chapter both focus on the discussion of the emotion extraction engine, which can detect and extract emotional feelings embedded in textual messages. The next chapter will move to another aspect of this research work: the generation of facial expressions. First the requirements of facial expression generation are discussed and then the system architecture is illustrated in detail. In the final part of the chapter, experiments assessing the performance of the synthesised images will be presented.

Chapter 6

Expressive image generator

The previous two chapters discussed the development of emotion extraction engine and a series of experiments that assessed the engine's performance. The emotion extraction engine can detect and extract emotional feelings from textual messages by applying linguistic and fuzzy logic knowledge. The experiments demonstrated that the emotion extraction engine can be used in the real-time communication context.

Online real-time expressive communication is beneficial as it provides some aspects of the visual clues that are present in face-to-face interaction but are not available in purely text based communications. One example of expression-enhanced communication is the use of emoticons [Windweaver 2003], which provide a sense of expression. Limitations of emoticons include the simplicity of expressions and the lack of intensity variations. With the rapid development of computer graphics, more and more researchers have focused on the creation of virtual reality (VR). Many recent works focused on creating an animated world or a realistic environment for Internet communication [Bryson and Levit 1991 and Vacca 1995]. One of the most important parts of VR is to create a virtualised or animated human or at least part of human, e.g., a talking head. Various synthetic facial models close to real human faces have been created [Badler et al. 1995, Badler et al. 2000, Rosis et al. 2000, Johnson et al. 2000, Hayes-Roth 2001, Hayes-Roth 2003, Conati 2002, Cassell 2000, Prendinger et al. 2003 and Rousseau and Hayes-Roth 1998]. However, applying realistic facial models for Internet communication requires significant computation time and large amount of data transferring that users with modest Internet connection speeds may feel reluctant to use.

It may take longer than expected for VR to emerge into normal Internet user's life and emotion icons or Internet avatars may be the dominant approach to express emotional feelings online for a long time. For this part of the research work, the specific problem addressed is how to generate the unique set of expressive images for human facial images and cartoon characters without the help of an artist. These generated expressive images can be used to present the emotional feelings of users in Internet communications (e.g. in Internet chat rooms) or for Internet avatars to present emotional feelings.

To generate facial expressive images in a fast method and in a user-friendly interface, hiring an artist to create professional drawings or letting users provide their own expressive photos would

be the easiest answer. The problems with both two approaches are obvious: If the system requires several expressions with various intensities, it would be unrealistic to force every user to supply 10-20 expressive images before he/she can login on the system. In an extreme case, when the system is applied in the Internet communication environment, it may attract thousands of potential users and it would be impossible for an artist to provide customised expression drawings to everyone.

As discussed above, manual methods (e.g. hire an artist or self drawing) can not fit in the Internet communication context. In this research work, an automatic method – the expressive image generator was created. This chapter first discusses the requirements of the expressive image generator (e.g. usage requirements, interface requirements and communication requirements etc.) and technologies applied in generating facial expressions, then the system architecture of the expressive image generator is explained in detail. In the final part of this chapter, two experiments assessing the recognition rates of the generated images are reviewed.

6.1 The requirements of the expressive image generator

The purpose of the expressive image generator was to synthesise facial expression images for the emotion extraction engine. As the emotion extraction engine requires images belonging to six categories and three intensities, a total of eighteen expressive images are needed to be synthesised by the expressive image generator.

6.1.1 Scenarios of the expressive image generator usage

The expressive image generator was designed to provide expressive images for the emotion extraction engine, which mainly targets the Internet communication context. The expressive image generator may also be applied in the Internet communication environment as a stand alone product such as providing real-time and dynamic facial expressions for an Internet game or generating facial expression animation for an online interactive movie. The expressive image generator should provide facilities that enable users to submit neutral facial images and guide the generator to synthesise facial images.

6.1.2 User interaction requirements of the expressive image generator

To synthesise facial expression images for Internet communication, the delay triggered by the expressive image generator should be kept a minimum and the user interaction with the generator should be clear and simple. Each user should only need to provide one neutral facial

image to the expressive image generator and users should have choices to undo and redo the synthesise operations to achieve a satisfactory facial expression.

6.1.3 Facial expression category requirements

In the Internet communication context, people may prefer to either remain anonymous or show their own identities. They may role play and change their genders, choose a nickname or use an image to symbolise themselves. Under some circumstances, they may prefer to use a cartoon image instead of their own images to represent themselves.

The expressive image generator is required to generate two types of image. One is human facial images and the other is cartoon facial images. Both types of image are widely used for Internet communication purposes (e.g. in Internet chat rooms and online games). Compared to human facial images, cartoon images have their own characteristics. First, the cartoon facial expressions can be extreme. Cartoons can show the extreme expressions and extreme behaviour without causing any negative feelings, e.g., the cartoon Simpson family's behaviour in shows. Second, parts of the characteristics presented in human images may be distorted or even disappeared in cartoon images, for example a cartoon image may have extremely large eyes and without eyebrows or upper lip.

The expressive image generator was designed to synthesise images belonging to the six universal expression categories, i.e., happiness, sadness, surprise, fear, disgust and anger. A wide range of expression intensity and variation of detailed expression exists within each category. The expressive image generator generates expressive images with three different intensities (low, middle and high) for each category. Facial expression animation is achieved in the expressive image generator as well.

6.1.4 The requirements of input/output and development language

Similar to the input and output requirements of the emotion extraction engine, interaction between the user and the image generator should also be simple. The number of steps required to operate the different features of the image generator should be kept to a minimum. Input into the system was designed to involve intensive use of the mouse. Output by the system to the user should be clear and attractive using popular multimedia formats (e.g. jpeg and gif). To develop the expressive image generator, the JAVA programming language was chosen as it is a platform independent and provides good low level image functions.

6.2 The development of the expressive image generator

In this study, two-dimensional facial expression models were chosen to decrease the data manipulation complexity. Instead of creating a facial expression template database, image warping and image morphing techniques were chosen as the foundation of the facial synthesis approach for this project.

To clarify the synthesis process, the following important prerequisites are made. The uploaded neutral human or cartoon facial expression image should be a frontal view only (no hats or glasses) and with a neutral expression. As well as the above requirements, uploaded neutral cartoon images need to contain a human-like face with similar expression features of humans (e.g. mouth, eye and eyebrows).

In contrast to the systems that use complex facial models and require a high degree of computation complexity to generate the synthesised images, the expressive image generator uses a relative simple algorithm based on image warping and image morphing. The facial model used by the expressive image generator requires users to identify only six control points and draw three control areas on the original neutral image. The expressive image generator will automatically generate 18 images corresponding to the six expression categories, each with three different intensities. These images are distributed among the users of the system and stored for future use, e.g., for the usage of the emotion extraction engine.

The working procedure of the expressive generator is described below. First, users upload a neutral facial image to the expressive image generator. The uploaded image can be either a human facial image or a cartoon face. Users need to indicate the type of the image by ticking the "human or cartoon" box. Second, users need to follow the on-screen instructions to identify the six start points and three control areas. Third, the synthesised facial images will be displayed on screen automatically. Users can choose whether to generate facial expression animations. If the answer is yes, short clips of facial animation will be created. The framework of the expressive image generator is illustrated in figure 31.

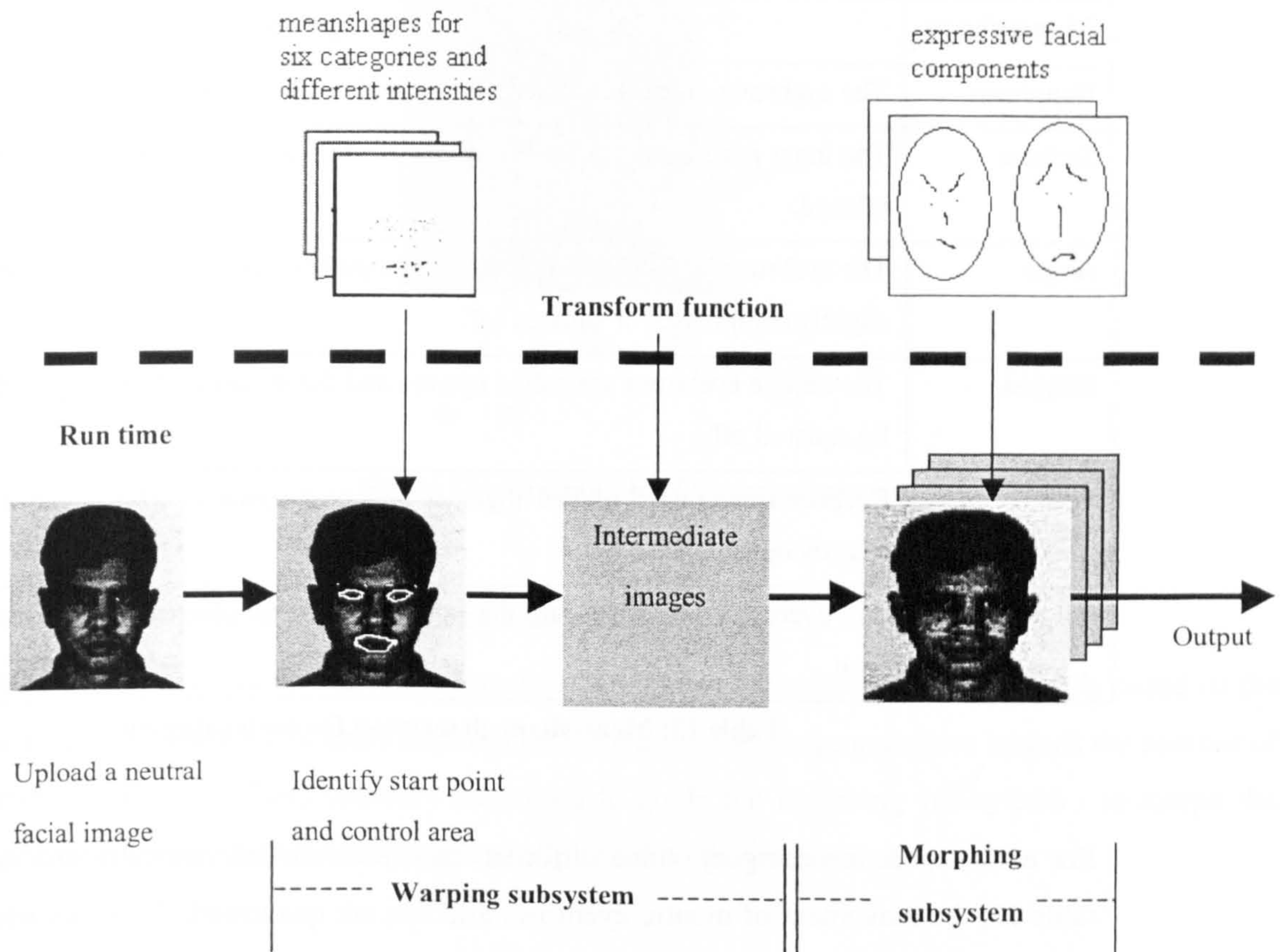


Figure 31: Expressive image generator framework

6.2.1 Expressive characteristics

In order to generate expressive images for the six expression categories, the characteristics of each category are evaluated. Ekman et al. [2003] and Parke and Waters [1996] demonstrated that 64 action units on the face control the generation of facial expressions. However determining the actual positions for the 64 action units require tedious interaction, which is not practical for real-time communication. In the expressive image generator, the action units for each expression type were simplified from Parke and Water's book [1996]. The simplified action units for each expression category are called the mean shapes for the corresponding category in this research work. The mean shapes for each expression category are listed in table 13.

Mean shape	Description
Happiness	The eyebrows are relaxed, the mouth is wide with the corners pulled up toward the ears.
Sadness	The inner portions of the eyebrows are piled up above the upper eyelid and the mouth is relaxed.
Anger	The eyebrows are pulled downward and together. The mouth is closed with the upper lip slightly compressed or squared off.
Disgust	The middle eyebrows are pulled upward and the mouth is slightly opened with the upper lip squared off.
Fear	Eyebrows are raised, pulled together and bent upward. The mouth may be dropped slightly open.
Surprise	The eyebrows are raised up, the upper eyelids are opened and the mouth is dropped open.

Table 13: Mean shape description for each category

For each expression category, three different intensities are automatically produced. Based on table 13, the movement of mouth, eyebrows and lips are quantified. As the expression intensity increases, the characteristics of corresponding expressions will also increase, which means that the movement progresses enlarged from low intensity to high intensity.

6.2.2 Image warping subsystem

6.2.2.1 Expression model mask definition

In the expressive image generator, the kernel that generates the images is named the “expression model mask”. The mask is constituted by two sets of points. The first set is the set of start points - R_1 selected by users. The second set is the control points of the start points, called the finish points - R_2 . The values of the finish points are calculated relative to the start points. By applying the masks to images of individual faces, corresponding expressions can be generated.

A general discussion about image warping has been presented in chapter 2. The expressive image generator takes the local area warping approach. Four segment areas are defined within the chosen region. The segments share a common apex R_1 . Other apexes depend on the area and the point R_2 . For example, to create a happy expression, the point R_1 can be chosen as the right edge of the mouth and the position of R_2 will be to the upper right. An example of segmented areas is shown in figure 32.

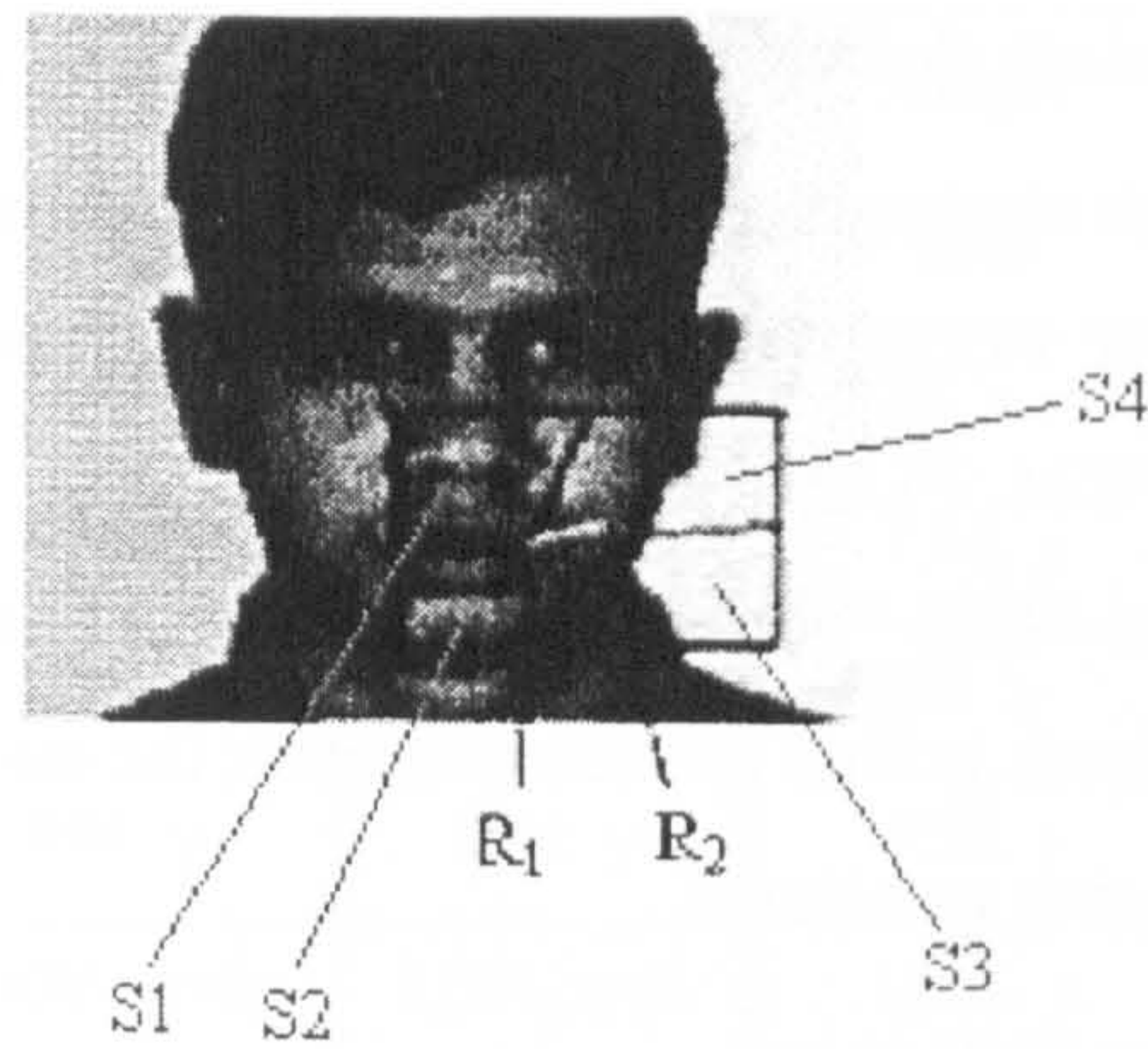


Figure 32: An example of the segment areas

6.2.2.2 Start point and control area collection

Users need to select the start points as the finish points are calculated automatically based on the start points. The control areas required by the expressive image generator include the contour of mouth and eyes. These contours are stored to guide the morphing sub-system - to morph the expressive mouth and eyes from image clips back to the intermediate image.

After uploading the neutral image to the expressive image generator, users will be guided to select six start points and three control areas. The six start points include: left corner of the mouth (LM), right corner of the mouth (RM), outer edge of the left eye (LOE), inner edge of the left eye (LIE), outer edge of the right eye (ROE) and inner edge of the right eye (RIE). The three control areas include the outer edge of the lips, and the inner edge of the eyelids for the left eye and right eye. An example of control points and control areas is shown in figure 33.

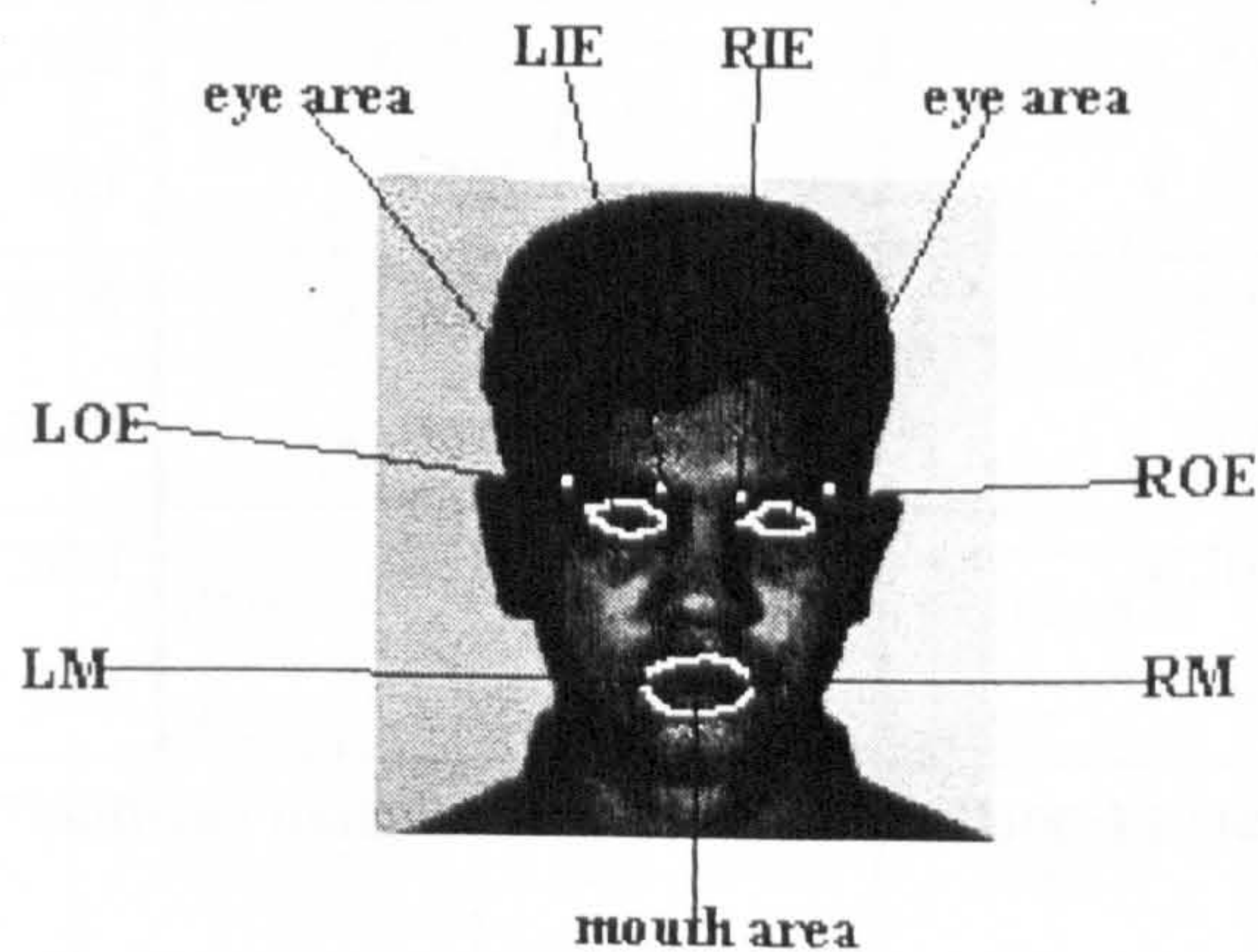


Figure 33: Example of control points and control areas

6.2.2.3 Finish points definition

In order to choose the segment areas, the finish points needs to be calculated. In total, six finish points FLM, FRM, FLOE, FLIE, FROE and FRIE are defined (FLM = finish point of LM, FRM = finish point of RM, FLOE = finish point of LOE, FLIE = finish point of LIE, FROE = finish point of ROE and FRIE = finish point of RIE). The functions to calculate the finish points are the same. The finish points are calculated as the start points plus an integer value that depends on the emotion being generated.

$$\text{Finish.x} = \text{Start.x} + a$$

$$\text{Finish.y} = \text{Start.y} + b$$

The actual values for all the finish points are shown in table 14-19. (Int1, Int2 and Int3 correspond to the intensity level of each category). Table 14-16 list the parameter values for human facial expressions and table 17-19 list the values for cartoon facial expressions.

	Happiness			Sadness		
	Int1	Int2	Int3	Int1	Int2	Int3
FLM. x	LM. x -2	LM. x-3	LM. x-4	LM. x-1	LM. x-3	LM. x-4
FLM. y	LM. y -2	LM. y-4	LM. y-6	LM. y+2	LM. y+3	LM. y+4
FRM. x	RM. x+2	RM. x+3	RM. x+4	RM. x+2	RM. x+3	RM. x+4
FRM. y	RM. y -2	RM. y-4	RM. y-6	RM. y+2	RM. y+4	RM. y+6
FLOE. x	LOE. x	LOE. x	LOE. x	LOE. x	LOE. x	LOE. x
FLOE. y	LOE. y	LOE. Y	LOE. y	LOE. y	LOE. y	LOE. y
FLIE. x	LIE. x	LIE. x	LIE. x	LIE. x +2	LIE. x+4	LIE. x+6
FLIE. y	LIE. y	LIE. Y	LIE. y	LIE. y-2	LIE. y-4	LIE. y-6
FROE. x	ROE. x	ROE. x	ROE. x	ROE. x	ROE. x	ROE. x
FROE. y	ROE. y	ROE. Y	ROE. y	ROE. y	ROE. y	ROE. y
FRIE. x	RIE. x	RIE. x	RIE. x	RIE. x	RIE. x-4	RIE. x-6
FRIE. y	RIE. y	RIE. Y	RIE. y	RIE. y	RIE. y-4	RIE. y-6

Table 14: Finish points value for human emotion "Happiness" and "Sadness"

	Surprise Int1	Surprise Int2	Surprise Int3	Anger Int1	Anger Int2	Anger Int3
FLM. x	LM. x	LM. x	LM. x	LM. x +2	LM. x+3	LM. x+4
FLM. y	LM. y	LM. y	LM. y	LM. y	LM. y	LM. y
FRM. x	RM. x	RM. x	RM. x	RM. x-2	RM. x-3	RM. x-4
FRM. y	RM. y	RM. y	RM. y	RM. y	RM. y	RM. y
FLOE. x	LOE. x+8	LOE. x+8	LOE. x+8	LOE. x	LOE. x	LOE. x
FLOE. y	LOE. y-3	LOE. y-6	LOE. y-8	LOE. y	LOE. y	LOE. y
FLIE. x	LIE. X	LIE. x	LIE. x	LIE. x +2	LIE. x+4	LIE. x+6
FLIE. y	LIE. y	LIE. y	LIE. y	LIE. y+2	LIE. y+4	LIE. y+6
FROE. x	ROE. x-8	ROE. x-8	ROE. x-8	ROE. x	ROE. x	ROE. x
FROE. y	ROE. y-3	ROE. y-6	ROE. y-8	ROE. y	ROE. y	ROE. y
FRIE. x	RIE. x	RIE. x	RIE. x	RIE. x-2	RIE. x-4	RIE. x-6
FRIE. y	RIE. y	RIE. y	RIE. y	RIE. y+2	RIE. y+4	RIE. y+6

Table 15: Finish points value for human emotion "Surprise" and "Anger"

	Disgust Int1	Disgust Int2	Disgust Int3	Fear Int1	Fear Int2	Fear Int3
FLM. x	LM. x-2	LM. x-4	LM. x-6	LM. x	LM. x	LM. x
FLM. y	LM. y+2	LM. y+4	LM. y+6	LM. y	LM. y	LM. y
FRM. X	RM. x+2	RM. x+4	RM. x+6	RM. x	RM. x	RM. x
FRM. y	RM. y+2	RM. Y+4	RM. y+6	RM. y	RM. y	RM. y
FLOE. x	LOE. x-0	LOE. x-0	LOE. x-0	LOE. x+8	LOE. x+8	LOE. x+8
FLOE. y	LOE. y-0	LOE. y-0	LOE. y-0	LOE. y-4	LOE. y-7	LOE. y-10
FLIE. x	LIE. x -8	LIE. x-8	LIE. x-8	LIE. x	LIE. x	LIE. x
FLIE. y	LIE. y-3	LIE. y-5	LIE. y-8	LIE. y	LIE. y	LIE. y
FROE. x	ROE. x	ROE. x	ROE. x	ROE. x-8	ROE. x-8	ROE. x-8
FROE. y	ROE. y	ROE. y	ROE. y	ROE. y-4	ROE. y-7	ROE. y-10
FRIE. x	RIE. x+8	RIE. x+8	RIE. x+8	RIE. x	RIE. x	RIE. x
FRIE. y	RIE. y-3	RIE. y-5	RIE. y-8	RIE. y	RIE. y	RIE. y

Table 16: Finish points value for human emotion "Disgust" and "Fear"

	Happiness	Happiness	Happiness	Sadness	Sadness	Sadness
	Int1	Int2	Int3	Int1	Int2	Int3
FLM. x	LM. x -6	LM. x-12	LM. x-18	LM. x-3	LM. x-6	LM. x-9
FLM. y	LM. y -5	LM. y-8	LM. y-10	LM. y+4	LM. y+7	LM. y+12
FRM. x	RM. x+6	RM. x-12	RM. x+18	RM. x+3	RM. x+6	RM. x+9
FRM. y	RM. y -5	RM. y-8	RM. y-10	RM. y+4	RM. y+7	RM. y+12
FLOE. x	LOE. x	LOE. x	LOE. x	LOE. x	LOE. x	LOE. x
FLOE. y	LOE. y	LOE. y	LOE. y	LOE. Y	LOE. y	LOE. y
FLIE. x	LIE. x	LIE. x	LIE. x	LIE. x +6	LIE. x+6	LIE. x+6
FLIE. y	LIE. y	LIE. y	LIE. y	LIE. y-4	LIE. y-8	LIE. y-14
FROE. x	ROE. x	ROE. .x	ROE. x	ROE. x-6	ROE. x-6	ROE. x-6
FROE. y	ROE. y	ROE. y	ROE. y	ROE. y-4	ROE. y-8	ROE. y-14
FRIE. x	RIE. x	RIE. x	RIE. x	RIE. x	RIE. x-4	RIE. x-6
FRIE. y	RIE. y	RIE. y	RIE. y	RIE. Y	RIE. y-4	RIE. y-6

Table 17: Finish points value for cartoon emotion "Happiness" and "Sadness"

	Surprise	Surprise	Surprise	Anger	Anger	Anger
	Int1	Int2	Int3	Int1	Int2	Int3
FLM. x	LM. x	LM. x	LM. x	LM. x+5	LM. x+10	LM. x+14
FLM. y	LM. y	LM. y	LM. y	LM. y	LM. y	LM. y
FRM. x	RM. x	RM. x	RM. x	RM. X-5	RM. x-10	RM. x-14
FRM. y	RM. y	RM. y	RM. y	RM. y	RM. y	RM. y
FLOE. x	LOE. x+8	LOE. x+8	LOE. x+8	LOE. x	LOE. x	LOE. x
FLOE. y	LOE. y-6	LOE. y-10	LOE. y-16	LOE. y	LOE. y	LOE. y
FLIE. x	LIE. x	LIE. x	LIE. x	LIE. x +2	LIE. x+4	LIE. x+6
FLIE. y	LIE. y	LIE. y	LIE. y	LIE. y+8	LIE. y+14	LIE. y+18
FROE. x	ROE. x-8	ROE. x-8	ROE. x-8	ROE. x	ROE. x	ROE. x
FROE. y	ROE. y-6	ROE. y-10	ROE. y-16	ROE. y	ROE. y	ROE. y
FRIE. x	RIE. x	RIE. x	RIE. x	RIE. x-2	RIE. x-4	RIE. X-6
FRIE. y	RIE. y	RIE. y	RIE. y	RIE. y+12	RIE. y+14	RIE. y+18

Table 18: Finish points value for cartoon emotion "Surprise" and "Anger"

	Disgust Int1	Disgust Int2	Disgust Int3	Fear Int1	Fear Int2	Fear Int3
FLM. x	LM. x-2	LM. x-4	LM. x-6	LM. x	LM. x	LM. x
FLM. y	LM. y+2	LM. y+4	LM. y+6	LM. y	LM. y	LM. y
FRM. x	RM. x+2	RM.x+4	RM. x+6	RM. x	RM. x	RM. x
FRM. y	RM. y+7	RM.y+14	RM. y+23	RM. y	RM. y	RM. y
FLOE. x	LOE. x	LOE. x	LOE. x	LOE. x+8	LOE. x+8	LOE. x+8
FLOE. y	LOE. y	LOE. y	LOE. y	LOE. y-4	LOE. y-10	LOE. y-16
FLIE. x	LIE. x -8	LIE. x-8	LIE. x-8	LIE. x	LIE. x	LIE. x
FLIE. y	LIE. y-6	LIE. y-10	LIE. y-18	LIE. y	LIE. y	LIE. y
FROE. x	ROE. x	ROE. x	ROE. x	ROE. x-8	ROE. x-8	ROE. x-8
FROE. y	ROE. y	ROE. y	ROE. y	ROE. y-4	ROE. y-10	ROE. y-16
FRIE. x	RIE. x+8	RIE. x+8	RIE. x+8	RIE. x	RIE. x	RIE. x
FRIE. y	RIE. y-6	RIE. y-10	RIE. y-18	RIE. y	RIE. y	RIE. y

Table 19: Finish points value for cartoon emotion "Disgust" and "Fear"

6.2.2.4 The transform function T

The segment areas are transformed to new segments by function T. The transform function is generated by the observation that the closer a pixel to the start point, the more important the pixel will be in the facial expression generation. When the segment areas are selected, the pixels in each segment will be transformed into the destination. As a result, some segments are stretched while others are suppressed, thus making facial expressions achievable. In figure 34, it can be seen that the segment areas 3 and 4 are compressed and areas 1 and 2 are stretched. By stretching areas 1 and 2 towards area 3 and 4, the lip is moved to upper right, which simulates the smile effect on the right hand side of the face.

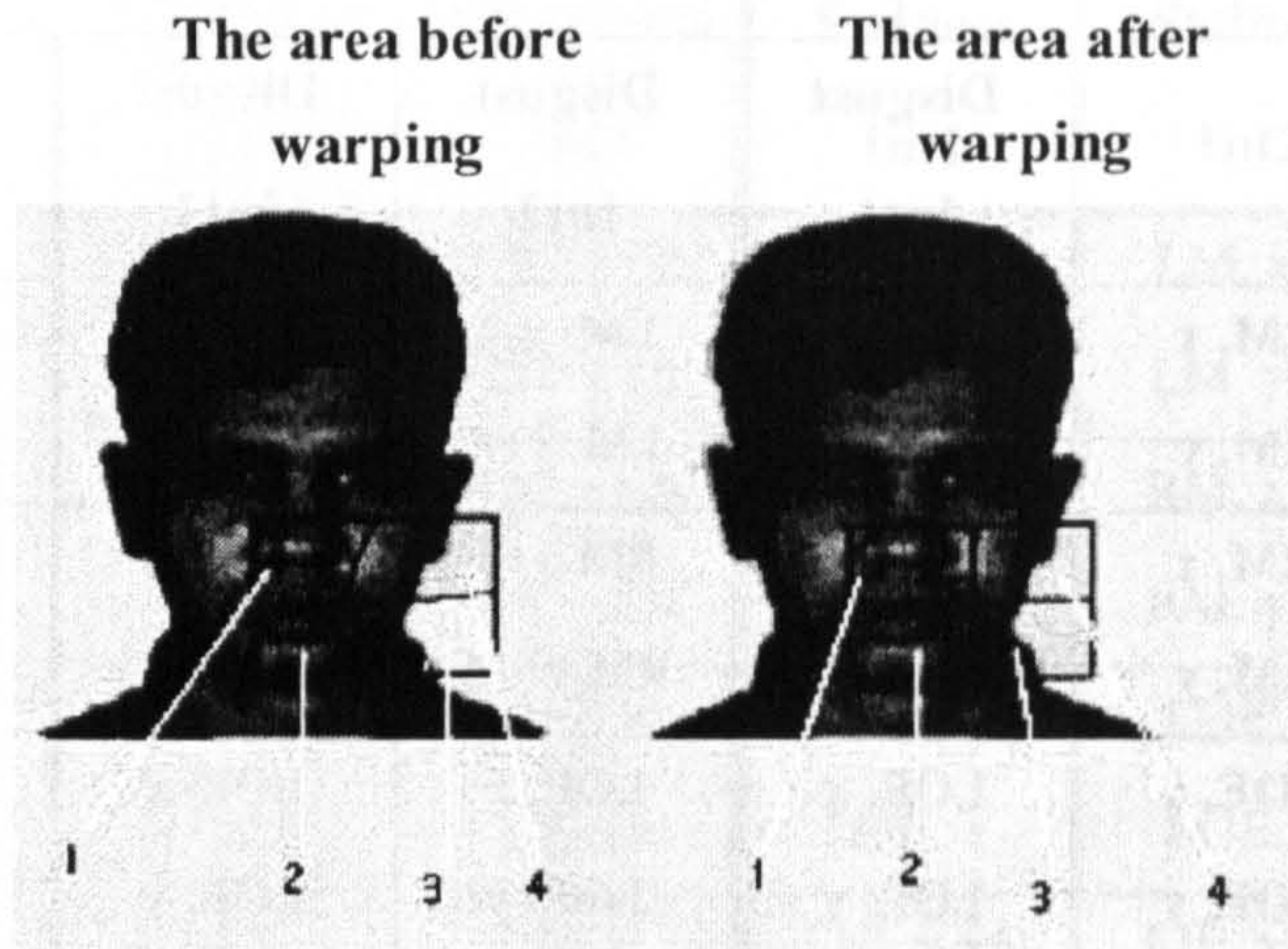


Figure 34: Local area warping example

Radial functions are applied in the transform function process. The characteristic feature of radial function is that the response decreases or increases monotonically with the distance from a central point, which suits the facial expression generation requirements. In the image generator, a typical radial function (Gaussian function with scalar input) is used to calculate the weight for each pixel. As each segment is scanned by x axis, an average weight for each line can be worked out. By dividing the weight for each pixel on the line by the average weight, the relative weight of each pixel can be calculated. The minimum relative weight is chosen to be 1 and the maximum relative weight is chosen to be 5 to reduce the possible gaps. The following formulas illustrate the calculations.

$$P(x_i) = \exp\left(\frac{|x_i - c|^2}{r^2}\right) \quad (6.1)$$

c is the central point (R_1) and r is the radius of the whole area $r = 2 \parallel R_1 - R_2 \parallel$

$$\text{Average} = \frac{\sum_{i=1}^n P(x_i)}{n} \quad (6.2)$$

n is the number of pixels in the scan line

$$N(i) = \begin{cases} 5 \\ \frac{P(x_i)}{\text{average}} & \text{when } 1 < \frac{P(x_i)}{\text{average}} < 5 \\ 1 \end{cases} \quad (6.3)$$

$N(i)$ is the relative weight of a pixel

The output of transform function $N(i)$ is the relative weight of each pixel. The relative weight determines the importance of each pixel, i.e. the number of times the pixel will be duplicated.

For example, if a segment is to be stretched, the pixels will be duplicated according to their relative weights. When a segment is to be compressed, the pixels with lowest weight will be removed.

For human facial images, the corresponding facial characteristics of the expression categories happiness, sadness, anger and disgust can be created by the image warping subsystem. For the expressions of fear and surprise, opened mouth and widened eyes are important characteristics that can not be achieved by the warping subsystem. To generate fear and surprise images, the intermediate images synthesised by the warping subsystem are sent to the image morphing subsystem.

For cartoon facial images, the characteristics of expressions sadness, disgust, fear and surprise are achieved by applying the morphing subsystem to the intermediate images generated by the warping subsystem. The expressions happiness and anger are directly generated by the image warping subsystem.

6.2.3 Image morphing subsystem

In the expressive image generator, image morphing is applied to the intermediate images generated by the image warping sub-system. In summary, some pre-selected expressive image clips (e.g. expressive mouths and eyes) are averaged and distorted to fit into the intermediated images.

The major difference of the characteristic between cartoon faces and human faces is the level of simplicity. Unlike ordinary humans, cartoons may not have a nose, a mouth or ears. The colours of human faces are extremely vivid; however there are usually fewer colours on a cartoon face.

According to the input images, the morphing subsystem can be divided into two types: the morphing system for human images and the morphing system for cartoon images. Both subsystems use numerous image clips of mouths and eyes that have previously been selected from different facial expression images. When the morphing subsystem receives the intermediate images, it will replace the mouth and eyes with the pre-selected mouths and eyes. The human morphing system will morph images belonging to expression categories fear and surprise and the cartoon morphing system will morph two extra categories: sadness and disgust to achieve relatively exaggerated expressions. In the cartoon morphing system, extra image clips such as tears and sweat are applied to generate the extreme expressions (e.g. crying with tears),

and an algorithm to change the size of eyes (e.g. to close the eye) is also implemented in the cartoon morphing system.

One common defect of morphing technology is the gap between the replaced part and the pre-selected images. To fill in the gaps that are generated where the mouth is opened or the eyes are widened, a Gaussian blur operation is applied. The structure of the human image morphing subsystem and cartoon image morphing subsystem are shown in figure 35 and 36.

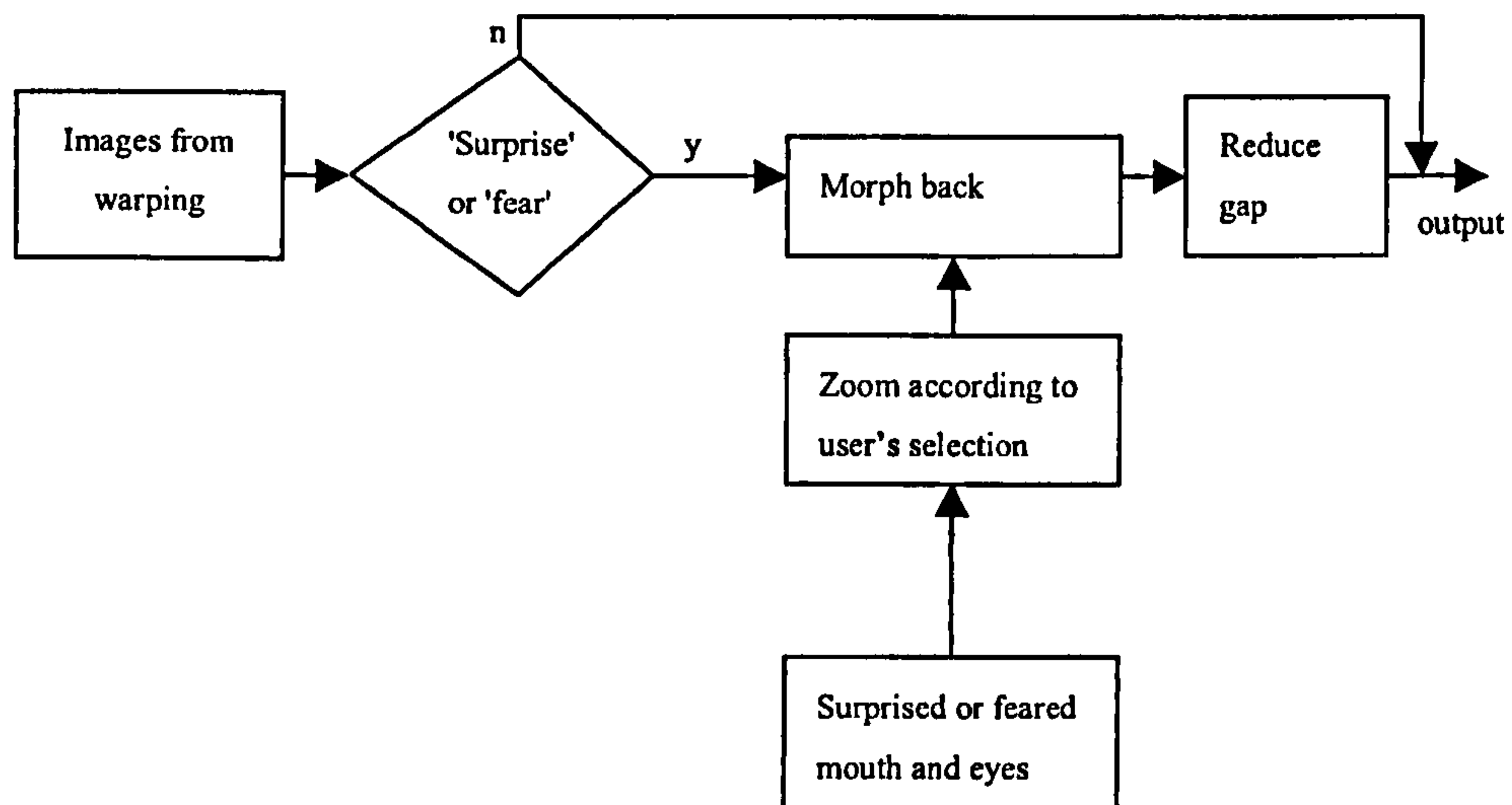


Figure 35: Structure of the morphing subsystem for human images

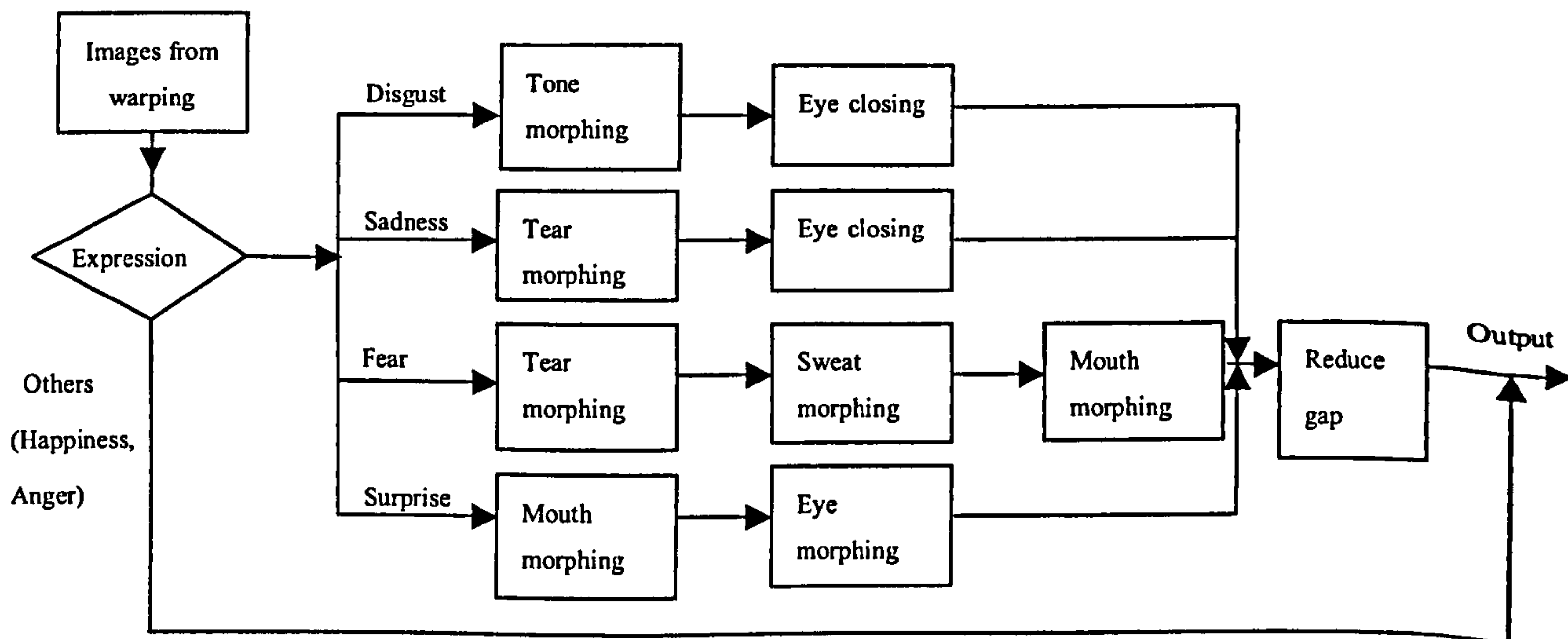


Figure 36: Structure of the morphing subsystem for cartoon images

Examples of expressive images generated by the expressive image generator are shown in figure 37 and 38. The default human image is a photograph of a colleague and the default cartoon image was obtained from the Metaface framework web site [Beard et al. 1999].



default image

Intensity increase from left to right



Emotion Happiness



Emotion Sadness



Emotion Surprise



Emotion Angry



Emotion Disgust



Emotion Fear

Figure 37: Generated human expression images

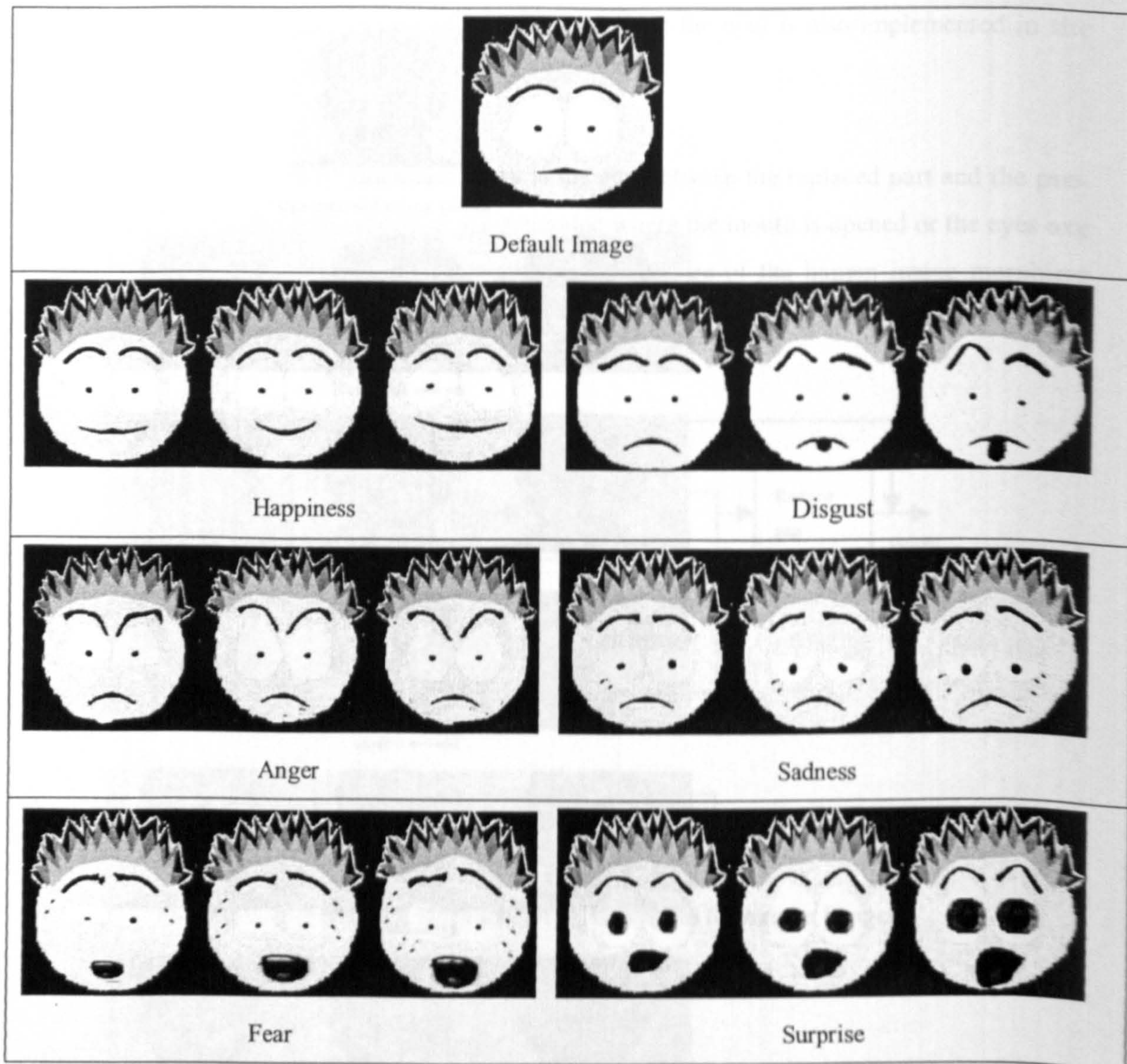


Figure 38: The cartoon expressive image set

6.2.4 Facial expression animation

The above discussions focus on the generation of discrete facial images. As the characteristics for each expression category and the parameter values for generating different intensities are already determined, synthesising expression animation can be easily achieved, e.g., by changing parameter values, using image morphing techniques or shape tween. In figure 39 and 40, facial expression animation clips generated by applying morphing techniques are presented to demonstrate the animation feature, where each line belongs to one individual expression category.

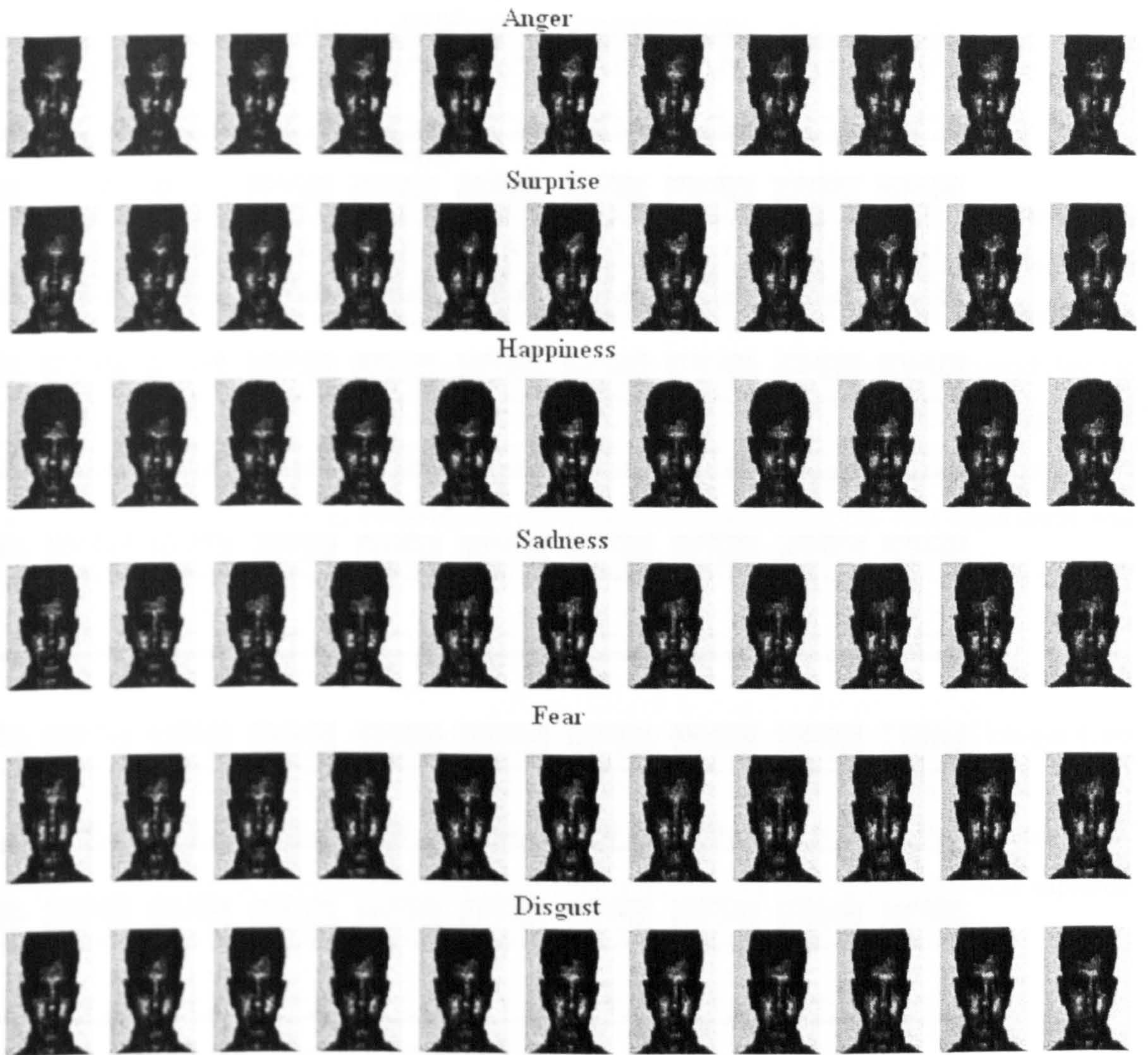


Figure 39: Human expression image clips



Figure 40: Cartoon expression image clips

6.2.5 Development summary

Above sections described the development of an expressive image generator. The expressive image generator generates not only a discrete set of expressive human images and cartoon images of varying expressive intensity, but also has the ability to synthesise continuous facial expressions. Although other researchers have developed a number of algorithms for facial expression generation, the problems of expression generation for Internet communication still exist. The major challenges are the computation complexity and interface complexity. The expressive image generator reduced the computation complexity by providing a warping and morphing model. Users only need to provide six start points and three control areas. Once the start points and control areas are selected, the expressive image generator can automatically generate facial expressions and it can be used in real-time applications. The following sections of this chapter will focus on the performance assessment of the expressive image generator and answer the most important question for the generator: Can the generated images be recognised by individuals?

6.3 Expressive image generator experiments

6.3.1 Planning the experiment

The expressive image generator synthesises a total of eighteen expressive images belonging to six expression categories and three different intensities for each neutral facial image. The developments of the expressive image generator lead to the following questions that need to be answered: Are the generated facial images (human and cartoon) recognisable by users? Are the expressions sufficiently recognisable by themselves, or do they have to have accompanying text? Two experiments were set up to answer these two questions. The main aims of the experiments were to determine whether users recognise the generated expressions. The first experiment was to assess the synthesised human facial images and the second experiment targeted cartoon images.

To assess the quality of the generated expressive images, four sets of expressive images (two human and two cartoon faces) were generated by the expressive image generator. The original cartoon neutral expression images were taken from the Internet site - Metaface [Beard et al. 1999]. One human face photograph was taken from a member of staff and another was captured from the Internet.

6.3.2 Method of the first experiment

For online communications, expressive images may be used in two different situations: standalone (e.g., online games or online chat rooms may show a user's expressions automatically) and text accompanied (e.g., users may occasionally type messages embedding emotions and an expressive image is displayed automatically).

To assess the recognition rate of the expressive images, two interfaces (image-only interface and image plus suggestive text interface) were created. The image-only interface contained the synthesised images in the six expression categories without any accompanying text. Users who viewed this presentation could only guess what expression the images presented from the images themselves. A question "What expression do you recognise from this image" was presented under each image. Users were asked to select from one of the six expression categories or choose "not sure" and type the name of the expression they identified.

The image plus text interface contained not only the synthesised images but also text that suggested the intended expression. The text contained information about the expression category of the image and the intensity of the image. For example, the accompanying text for the generated sad expressive image was "I am sad". A question "Compared to the default neutral image, can you recognise a sad expression from the corresponding picture?" was presented. Users were required to select "agree", "not sure", or "not agree".

- **Experiment operation**

A total of 35 participants were selected randomly from the students and staff of Bournemouth University. Each participant viewed both the image-only interface and image plus text interface. The viewing sequences of the different types of expressions and intensities were evenly distributed among the participants. The experimenter told each participant that she or he would watch two interfaces that contained expressive images belonging to six expression categories with different expressive intensities. After the viewing, the participants were asked to answer the questions shown on the interface. The experimenter also explained the purposes of the experiments and the operational procedures for the interfaces.

- **How performance was measured and analysed**

For both interfaces (the image-only interface and image plus text interface), the measurement was the number of expressive images that participants correctly identified. This was measured by analysing the participants' answers to the question "What expression does this image describe" and "Compared to the default neutral image, can you recognise the expression from the corresponding picture".

- **Result analysis**

The results of the experiment are shown in table 20 and 21. The data in both tables represents the percentage of participants who correctly identified the expressions. Table 20 shows the percentage in the "image only" test and percentage in the "image plus text" are shown in table 21.

	Happiness	Sadness	Surprise	Disgust	Anger	Fear
Low Intensity	0.34	0.74	0.56	0.1	0.66	0.22
Middle Intensity	0.68	0.89	0.73	0.45	0.73	0.61
High Intensity	0.92	0.91	0.87	0.61	0.76	0.73

Table 20: The percentage of the participants who correctly recognised human expressions in image only interface

	Happiness	Sadness	Surprise	Disgust	Anger	Fear
Low Intensity	0.47	0.83	0.72	0.32	0.72	0.3
Middle Intensity	0.83	0.9	0.83	0.78	0.78	0.76
High Intensity	0.95	0.9	0.92	0.84	0.9	0.93

Table 21: The percentage of the participants who correctly recognised human expressions in image plus text interface

It is shown that by accompanying the expressive images with text, the participants correctly recognised more expressive images than in the image-only test. By increasing intensity, even more participants also correctly recognised the expressive images. For the expression categories of happiness, sadness, surprise and anger, more than 70% of the images were correctly recognised when the intensity level is middle or high. For the expression categories of fear and disgust, on average more than 60% of the images were correctly recognised when the intensity is middle or high.

The chi-square test was carried out to test whether participants chose their answers randomly. The results showed that for the six expression categories happiness, sadness, surprise, anger, fear and disgust, the corresponding obtained chi-square values are 245.3, 456.1, 563.2, 123.4, 245.7 and 156. The calculated level of significance is less than 0.01 ($p=0.009$). As a result, the null hypothesis, that the experiment participants chose their answers randomly can be rejected.

6.3.3 Method of the second experiment

The main aim of this experiment was to determine whether users can recognise the synthesised cartoon expressive images. The experiment process is similar to the above experiment. Fifty individuals (not the same experiment participants in the human image tests) participated in the experiment. In a similar manner to the first experiment, the participants viewed two cartoon

expression interfaces (with accompanying text and without) and then answer the questions that were asked. The percentages of the participants who correctly recognised the cartoon expressions in the image-only interface are shown in table 22 and the percentages of the participants who correctly recognised the cartoon expressions in the image plus text interface are shown in table 23.

	Happiness	Sadness	Surprise	Disgust	Anger	Fear
Low Intensity	0.74	0.89	0.81	0.59	0.75	0.63
Middle Intensity	0.8	0.94	0.73	0.72	0.78	0.68
High Intensity	0.68	0.9	0.73	0.62	0.75	0.79

Table 22: The percentage of the participants who correctly recognised cartoon expressions in image only interface

	Happiness	Sadness	Surprise	Disgust	Anger	Fear
Low Intensity	0.78	0.93	0.81	0.79	0.84	0.86
Middle Intensity	0.88	0.92	0.79	0.85	0.87	0.86
High Intensity	0.75	0.9	0.81	0.68	0.7	0.83

Table 23: The percentage of the participants who correctly recognised cartoon expressions in image plus text interface

It is shown that more than 59% of experiment participants correctly guessed the expression category in the image-only interface and more than 78% of experiment participants identified the correct category in the image-plus-text interface.

The chi-square test was carried out on the results of the cartoon image experiment. For the six expression categories, the corresponding obtained chi-square values are 123.9, 235.7, 178, 245, 148 and 230. The calculated level of significance was less than 0.01 ($p=0.005$). As a result, the null hypothesis that experiment participants chose their answers randomly in the cartoon image test can be rejected.

6.3.4 Experiment summary

From the results of the two experiments, the following conclusions may be drawn.

- For human facial images, the recognition rate improves as the intensity increases.
- For cartoon facial images, the images with middle intensity were the most successfully identified.

For human expression images, low intensity images are not easily recognised by users, however as the intensity increases, the recognition rate can be improved. For cartoon expression images, middle intensity images achieved the best recognition rate.

For human expression images, the recognition rate varies. For cartoon expression images, on average more than 65% of the images were recognised correctly. For the expression categories of happiness, sadness, anger, fear and surprise, on average more than 75% of the images were correctly recognised for their intended purpose by the participants. It is shown that by the addition of text to both human and cartoon expressive images, the participants correctly recognise more expressive images than without the text.

The human image experiment showed that as intensity increased, the participants could recognise more expressive images. The cartoon image experiments show that by increasing the expression intensity up to the middle intensity, more recognisable expressive images can be obtained.

By comparing the recognition rate of the cartoon expression images with the human expression images, it is found that for both human facial images and cartoon images, expression intensities do influence the users' perceptions. As intensity increases, the expressive characteristics become clearer, and the users' perception ability is increased. The cartoon image generator intentionally exaggerated the expressive features. Although the exaggerations do improve the user's perceptions, for a high intensity expressive image, too much exaggeration may cause some negative effects: some users who correctly identified the middle intensity images may identify high intensity images incorrectly.

6.4 Chapter conclusions

This chapter described the development of an expressive image generator for both human and cartoon images. The expressive image generator can synthesise not only discrete set of expressive human and cartoon images of varying expressive intensity but also expression animations. To fit for the target context – Internet communication, the expressive image

generator reduces computation complexity and interface complexity by using image warping and image morphing.

The experiments demonstrated that more recognisable expressive images could be obtained by increasing the intensity. For human expression images, the recognition rate increased as the image intensity increases, while for cartoon expression images, the recognition rate declined beyond the middle intensity.

The experiment results also proved that the generator could create relatively satisfactory expressive images suitable for an on-line chat environment. The rates of recognition increased when accompanied by supporting text. The rates of recognition are acceptable for informal leisure pursuit applications in which accompanying text is present.

Up to this chapter, the emotion extraction engine which extract emotional feelings from textual messages and the expressive image generator which generates facial expressions for both human and cartoon images have been discussed. The two systems aim at answering the questions of “how to detect and extract emotional feelings from textual messages?” and “how to display the emotional feelings extracted?”. The frameworks to solving these questions have been discussed from chapter 4 to this chapter – chapter 6. The next chapter will focus on the applications of the emotion extraction engine. Four applications –a stock market emotion analyser, a 2D expressive chat interface and a 3D expressive chat interface will be presented.

Chapter 7

Applications of the emotion extraction engine

The emotion extraction engine detects and extracts emotional feelings from textual messages. The expressive image generator synthesises expressive images, which can be used by the emotion extraction engine. To demonstrate the usability of the emotion extraction engine, a series of applications of the emotion extraction engine were developed. In this chapter, a stock market emotion analyser, a 2D chat application and a 3D chat application are discussed.

7.1 Stock market emotion analyser: tailoring the emotion extraction engine

Emotional feelings can be observed not only in human to human communication or human-machine communication, but also in other interaction styles. One of the most interesting examples is the stock market. Although frequently used in papers and articles, the term stock market emotion does not have an authoritative definition as this term is largely influenced by non-scientific use and meaning. Stock market emotions are defined in this research work as "the different states of expectation investors derived from their perception of trends in the movement of share prices". The stock market emotions include three different states, i.e. happiness, sadness and stable.

The state happiness (an 'up' emotion) matches in the situation where share values are rising. The state sadness occurs when share values are falling (a 'down' emotion). If there is no change in the value, the market will be in the state stable. From this definition, the stock market emotion directly reflects the movement of the stock price index. In chapter 2, a simple stock and user interaction model is presented and here a summary of the model is described in the following paragraph for convenience.

The stock market and shareholder interaction process can be described like this: shareholders buy or sell shares in the stock market. The stock market generates data such as the share value and the volume of trading. Stock analysis experts analyse the data and write articles describing it. TV, radio or newspapers present these articles to the public. The articles may be read by the

shareholders and my influence their future buying or selling behaviours, which will restart the loop.

By applying the emotion extraction engine to stock market articles, a stock market emotion analyser can be created. The stock market emotion analyser can detect the embedded emotional feelings from stock market articles. To detect the relationships between the stock market emotions and the future behaviour of the investors, first, the characteristics of the stock market analysis articles were examined and then a series of experiments were carried out.

7.1.1 Stock market analysis articles

In general, the articles can be classified into two categories: objective-dominated articles and opinion-dominated articles. Examples of an objective-dominated article and an opinion-dominated article are shown in figure 41 and 42.

*"The FTSE 100 remained in negative territory on Friday in light volume after the Dow extended its slide after the London close.
The London blue-chip index traded down 0.3 per cent at 3,674.5 at 1000 GMT while the mid cap FTSE 250 index continued to trade in muted fashion, down 0.2 per cent at 4,062.9.
The Dow Jones Industrial Average closed down 85.64, or 1.1 per cent, at 7,914.96 reacting negatively to a record trade deficit and a leap in producer prices. But a bullish research note from Merrill Lynch on semiconductors buoyed the tech-laden Nasdaq Composite, which eased just 3.09, or 0.2 per cent, to 1,331.23. "*

Figure 41: Objective-dominated article

*A Competition Commission inquiry into extended warranties is unlikely to find that Dixons, one of the UK's largest electrical goods suppliers, has a monopoly in the market.
The Commission - which is probing claims of overpricing and lack of choice in the extended warranty market - said it had reconsidered its view in January that Dixons may have a "scale monopoly" - or more than 25 per cent of market share.
"Further analysis, taking account in particular of the supply of free warranties, has, however, suggested that Dixons share is slightly below 25 per cent," the Commission said.*

Figure 42: Opinion-dominated article

Opinion-dominated articles contain more analyses than the objective-dominated articles and may influence the reader's judgement stronger than the objective-dominated articles. However, the problem with opinion-dominated articles is bias - two opposite analysis results can be based on the same fact. In this research work, the articles input into the stock market emotion analyser are chosen to be objective-dominated articles to minimise the influence of subjective opinions.

7.1.2 Emotion definition in the stock analyser

In the previous section, three emotion states of the stock emotion were defined. In the stock market analyser, the emotion state stable was not implemented. One reason is the absolute stable state is quite rare. The other reason is that the stable state may be represented by state happiness and sadness. For example, if the total number of happiness and sadness is equal, then that state may be described as stable.

The stock market emotions in the analyser are defined in two levels: the sentence analysis level and the article level. For the sentence analysis level, the analyser will consider a sentence as happiness only if the sentence describes an up trend. For example "the overall value is 5% up" will be considered into emotion happiness. The opposite condition applies for the emotion sadness. In the article level, the analyser will consider the overall emotion of an article as the difference between the numbers of the detected emotion states in all the sentences. If more happiness emotion is found, the article's state will be happiness, otherwise it will be sadness.

For both emotion states, different intensities exist. In the sentence analysis level, the intensity is defined as the absolute difference between the numbers of up and down emotions in the sentence. In article level, the emotion intensity is defined as the absolute difference between the numbers of up and down sentences.

7.1.3 Emotion measurement

Tools and research prototypes, e.g., Arps index [Wood and Arps 1996] and CBOE index [Whaley 2002], have been developed to measure the stock market emotions. These tools are based on the analysis of market values and column data. For example, The CBOE Market Volatility index (VIX) is calculated in real-time using the Standard and Poor's 100 Index (OEX) options.

For the stock market emotion analyser, the input is the stock market news instead of the real market trading data. To measure the emotions in the articles, different weights and different categories are assigned to different emotional words (in the analyser, the weights are called the intensities). For example, word "up" belongs to emotion category happiness and is assigned intensity 1. The word "large" is assigned an intensity of 2. In this case, the sentence "there is an up movement" is treated as a sentence containing emotion happiness of intensity 1, and sentence "there is a large up movement" is classified as a sentence containing emotion happiness but with intensity 3.

7.1.4 Stock market emotion analyser architecture

The same as in the emotion extraction engine, the stock market emotion analyser contains an interface layer and emotion analysis components. The analysis components include three parts: the input analysis, the tagging system and the parser.

- **The input analysis**

The main duty of the input analysis function is the same as the emotion extraction engine's input analysis function - to divide the articles into sentences correctly. However, the input analysis function is extended in the stock market emotion analyser as some challenging special characters occur more often in stock market articles than in other contexts. For example, a dot can be not only a terminator, but also a decimal point, e.g., "The index was up 10.3 per cent." Terms such as "10.3" are widely used in stock market articles. To correctly understand the usage of dot, an analysis was carried out to test whether it is a terminator or is used for other purposes.

- **The tagging system**

Most tag-sets the stock market emotion analyser used are remained the same as in the emotion extraction engine, except for the following tags. First, in the emotion extraction engine, the tag "EMO_W" is assigned to each possible emotion word. In the stock market emotion analyser, the tag "EMO_W" is replaced by a new tag "STO_W" (stock word), which is assigned to each possible emotion word related to the stock market. Second, in the emotion extraction engine, there are six emotion categories and three intensities. For the stock market emotion analyser, only two emotion categories (happiness and sadness) are defined. Accordingly, the tag-set uses the numbers 0 to 1 to represent happiness and sadness. Third, although the stock market emotion analyser still tag the entire word in order to keep the minimum response time, the possible emotion word and its related intensity are treated differently compared to the tagging

system in the emotion extraction engine. For example, the word "negative" in the stock market context is an emotion word and its category is sadness. In human-human communications, the word "negative" does not always carry direct emotional feelings. A subset of the tag-set used in the stock market emotion analyser is shown in table 24.

Tag	Description	Example
NOUN	Noun	Kindness
1N2	Happiness, Noun, intensity 2	Happy
0N2	Sadness, Noun, intensity 2	Sad
PC	Present continuous	Is having
PP	Present perfect	Have seen
NDP	negative data point	End
NT	Noun third person	She

Table 24: The tag subset for the stock analyser

- **The tagged dictionary**

The tagged dictionary was extended from the one used by the emotion extraction engine, but the structure remained the same. To find the possible emotional words, the financial newspapers such as The Financial Times were manually examined and the contained emotional words were extracted and then added to the dictionary. An example of the dictionary entries is shown in table 25.

Word	Word category	Emotional tag
Happy	1N2	STO_W
She	NT	

Table 25: Tagged dictionary example

- **The parser**

The parser's working flows are kept the same as the emotion extraction engine. The output of the parser is still the emotion category, the intensity and the tense - [emotional category] [intensity] [tense]. In this section, several sentence parsing examples are listed. For example, when no emotion word is detected in a sentence, the parser will treat the sentence as a neutral emotion sentence. When conditional emotions are detected in a sentence, e.g., *'the stock market will be stronger if the war finishes quickly.'*, the parser will interpret this sentence as an emotional sentence with decreased intensity. When more than one emotion categories' words

are detected in a sentence (e.g. sentence such as *'The FTSE 100 was down but FTSE 500 was up'*), then the parser will treat the sentence as an emotional sentence with two states. The parser will not apply methods to resolve conflicting emotions because the stock market articles may just want to express a mixed feeling.

- **The interface layer**

Except for interacting with users, the main duty of the interface layer is to display the overall emotion status according to each sentence's emotion and intensity. The interface shows the overall emotion status of the article and the overall intensities of emotion happiness and sadness. An example screen of the interface is shown in figure 43.

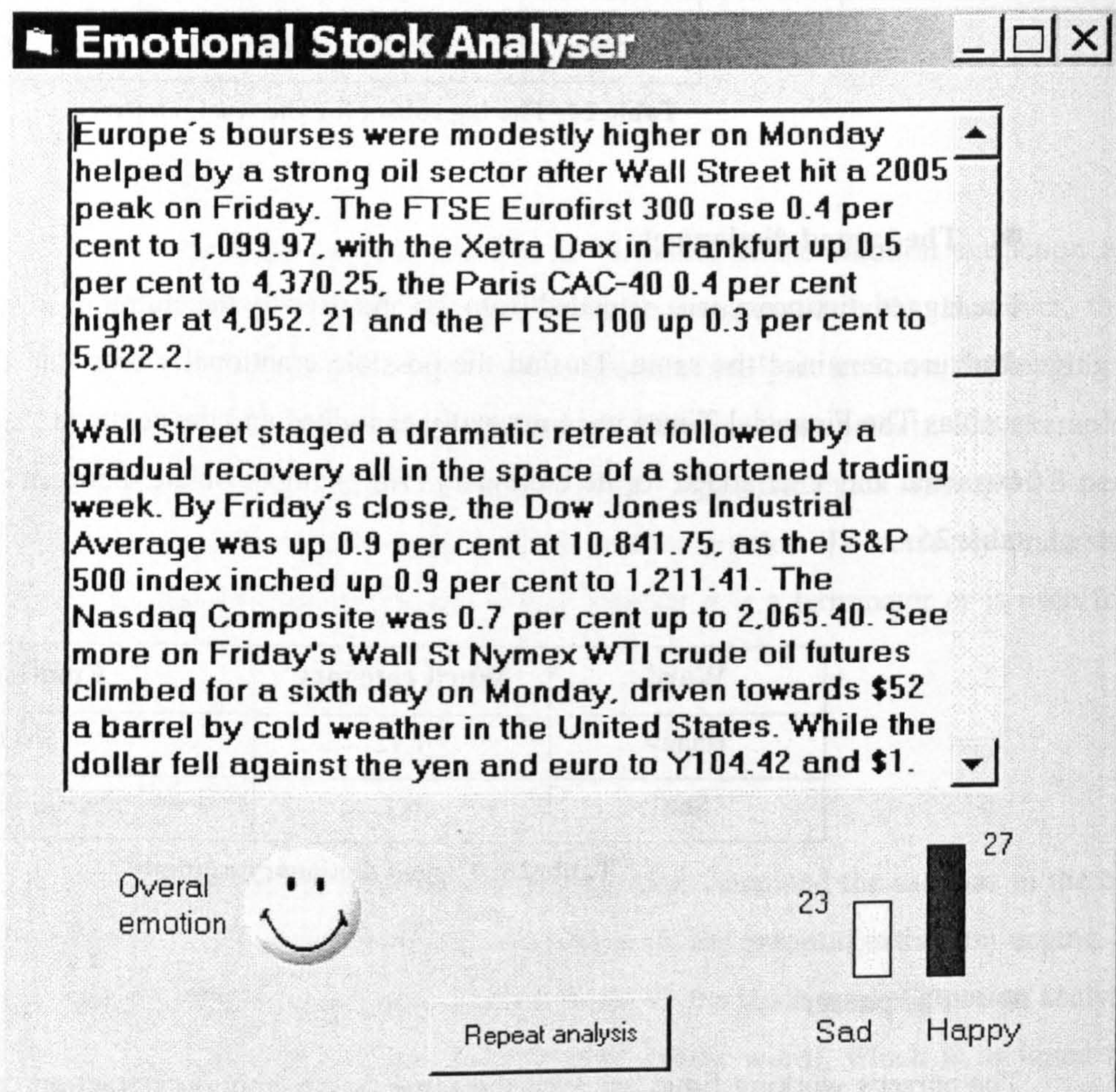


Figure 43: The interface of the analyser

7.1.5 Stock market analyser experiment

The stock market analyser is designed to detect the emotions embedded in stock market articles. As the stock market analyser adopts the emotion extraction engine and the engine had been

tested, experiments similar to previous tests were not conducted. Instead, the correlation between the output of the stock market analyser - emotion status and the future movement of the stock market were tested. An experiment aimed at assessing the possible correlations between the emotional feelings and the stock price movements was planned.

7.1.5.1 Methods of the experiment

The aim of the experiment was to determine whether there was any correlation between the detected emotion status of the articles and the change in value of the London stock market. The emotional status of the articles is compared with FTSE (Financial Times Stock Equity) -ALL index values and FTSE 100 values.

Fifty articles were chosen from the Financial Times and its web site, representing one article per week from the period 26-04-2002 to 27-03-2003. The chosen articles were purely about that day's overall stock market index data with a minimum of personal opinions. As a sufficient amount of articles about individual industry sectors or companies were not found, the experiment only assessed the overall market level.

7.1.5.2 Percentage analysis

In the following figures, the "correct" section contains the results where there was a consistency with the output of the analyser and the movement of the stock market index. When the analyser's output indicates "sad" and the next day's FTSE-ALL index or FTSE 100 is down, and when analyser's output indicates "happiness" and the next day's FTSE-ALL index or FTSE 100 is up. The "not correct" section contains the results where there was no consistency with the output of the analyser and the movement of the stock market index. The FTSE-ALL index result and FTSE 100 are shown in figure 44 and 45.

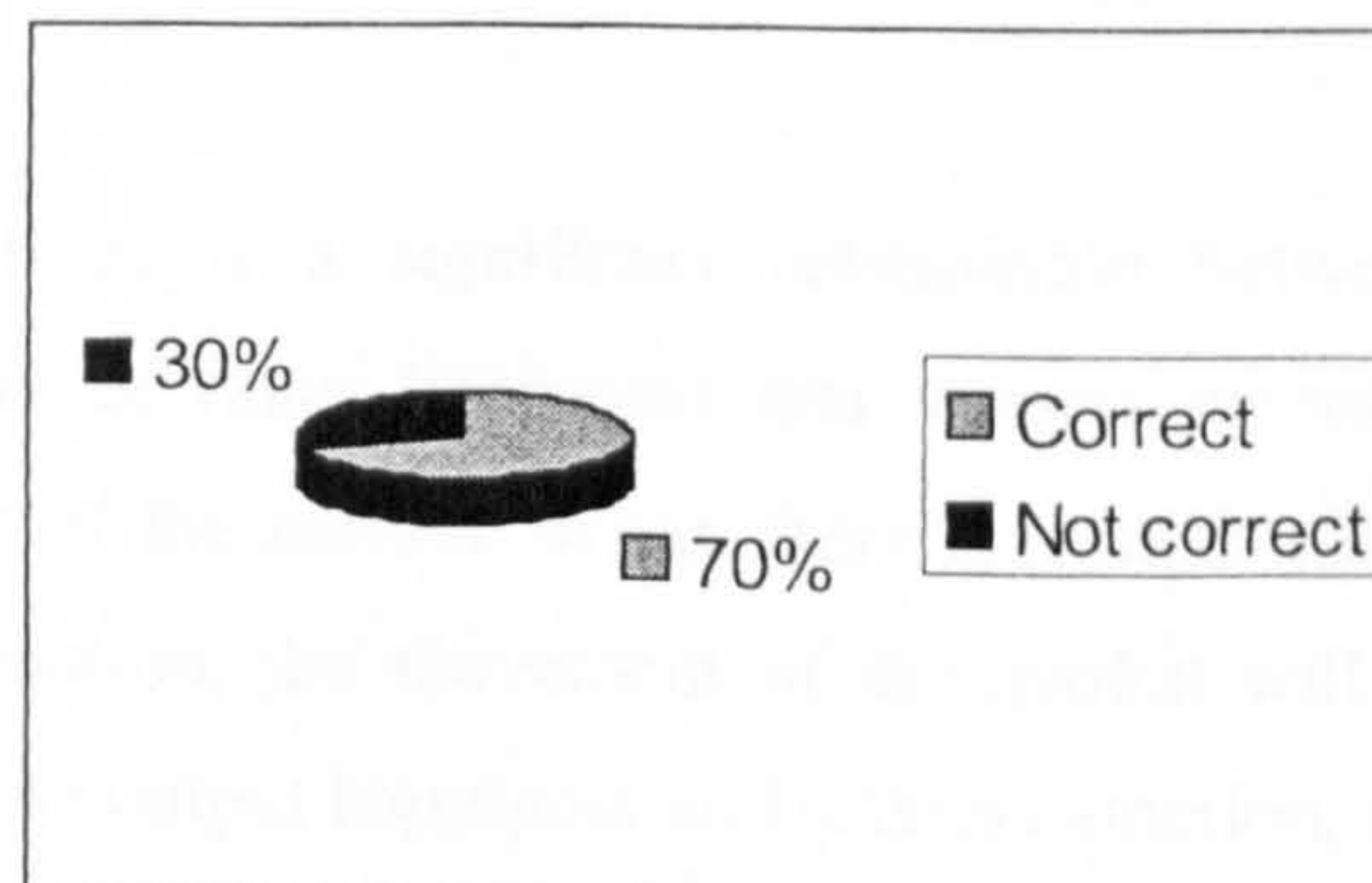


Figure 44: The FTSE-ALL index percentage results

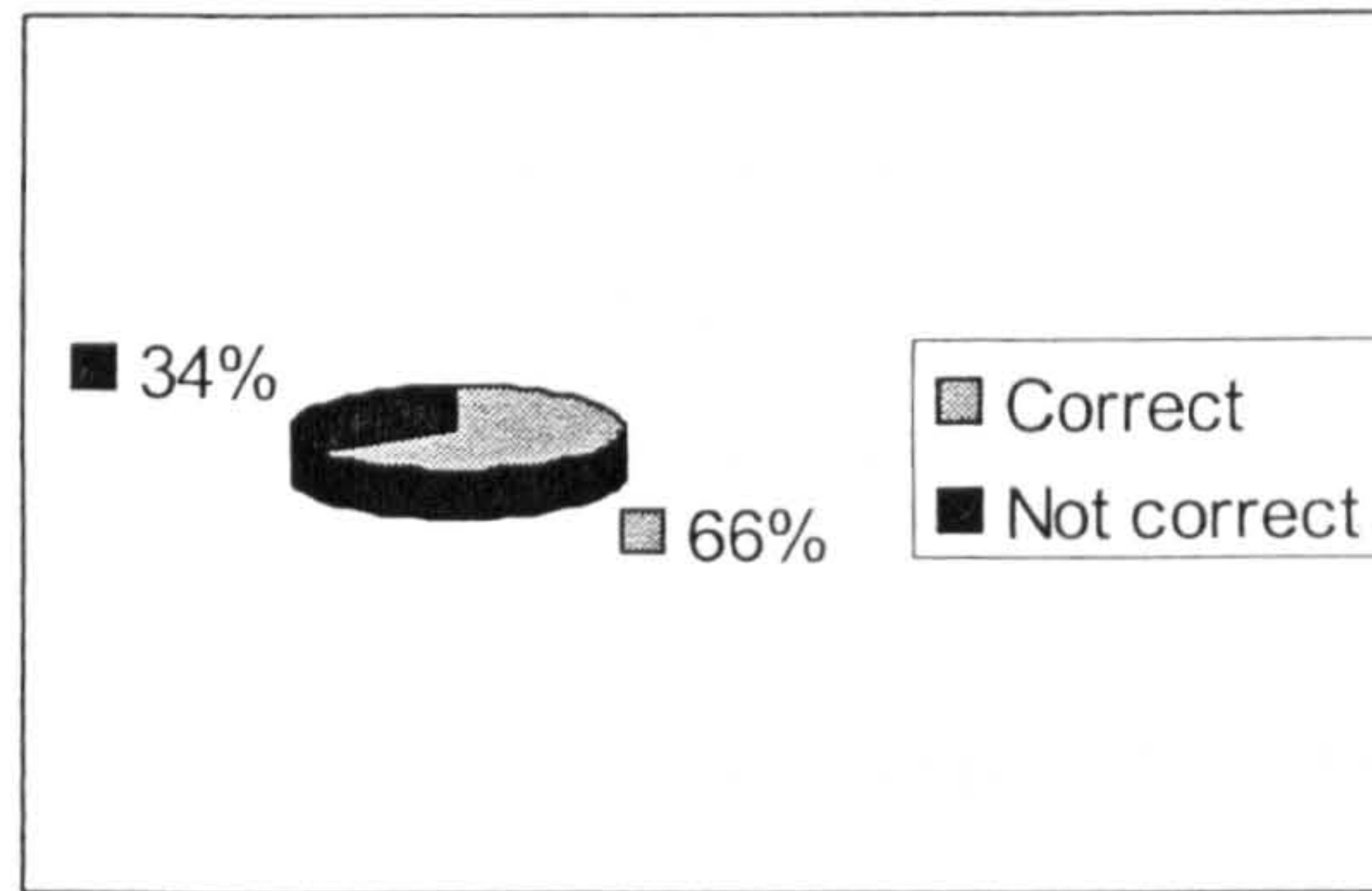


Figure 45: The FTSE100 percentage results

7.1.5.3 Correlation analysis

Correlation analyses were performed to determine whether significant relationship exist between the emotions detected in the articles and the down or up movement of the market the next day. For example, if the stock market emotion analyser found 20 happiness and 40 sadness emotions, will next day's share value decline be heavier than the day that 10 happiness and 13 sadness are found?

Test 1: overall data analysis

The absolute values of the difference between happiness and sadness emotions is defined as DIFFER: $\text{DIFFER} = |a-b|$, where a is the total happiness intensity and b is the total sadness intensity.

- **FTSE all-index values**

The DIFFER and the next day's absolute index movement were found to be significantly related. The correlation result is 0.284, which is significant ($p=0.05$). The correlation results are shown in figure 46.

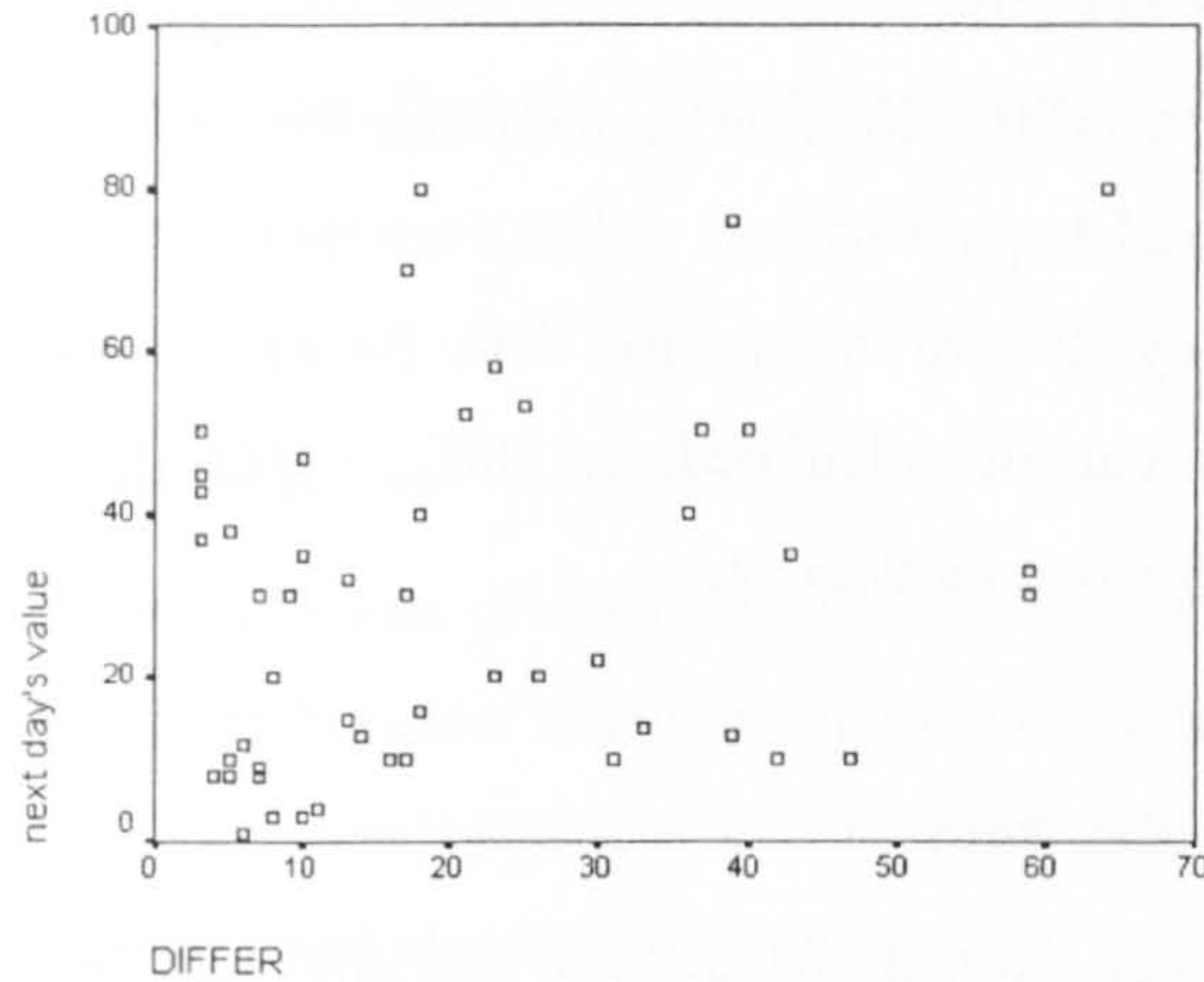


Figure 46: FTSE-ALL index correlation between DIFFER and next day's value

● FTSE 100 data

A marginally significant correlation between the DIFFER and next day's absolute index movement was found. The correlation result is 0.259, which is marginally significant ($p=0.07$). The results are shown in figure 47.

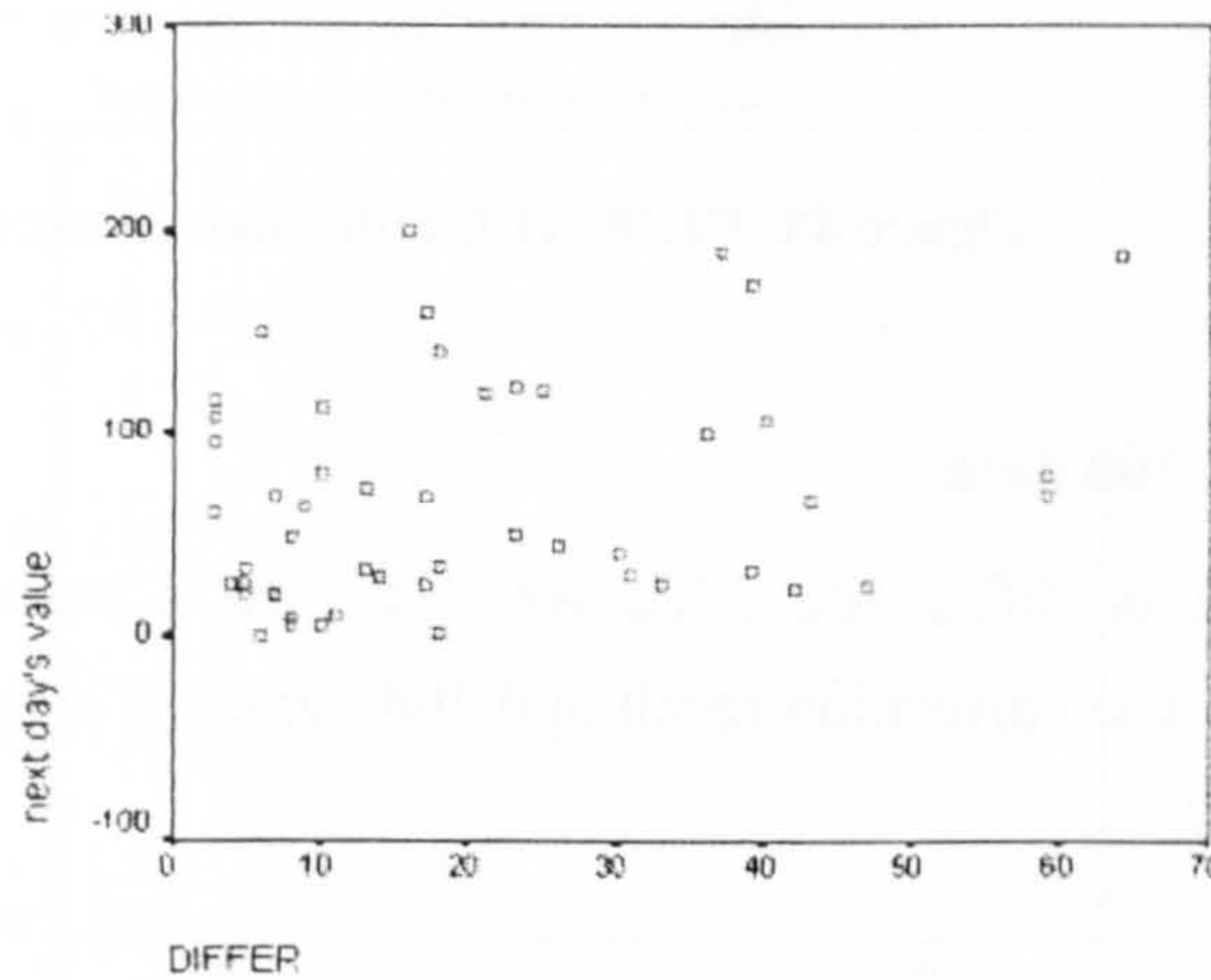


Figure 47: FTSE 100 correlation between DIFFER and next day's value

The results show that there is a significant relationship between DIFFER (the absolute differences of the number of times happiness and sadness emotions are presented) and the absolute index movement of the market. When there is a small difference between the output happiness and sadness emotion, the movement of the market will be small. When there is a large difference between the output happiness and sadness emotion, the movement of the market will be large. However no significant correlation between the direction of the market movement and the detected emotion difference was detected.

Test 2: movement analysis

In this test, four terms diff_1 , diff_2 , diff_{ew} and diff_{sw} were defined. Diff_1 is the intensity difference between the detected happiness and sadness emotion at a specific date. Diff_2 is the intensity difference seven days before the specific date. By minus diff_1 by Diff_2 , the consecutive weekly emotion difference can be calculated, i.e. $\text{diff}_{ew} = \text{diff}_1 - \text{diff}_2$. diff_{sw} is defined as the actually consecutive weekly stock movement.

- **FTSE-ALL index data**

Correlation tests were carried out between the diff_{ew} and diff_{sw} , and the test result was 0.365 which shows that a significant correlation ($p=0.01$) exists. The results are shown in figure 48.

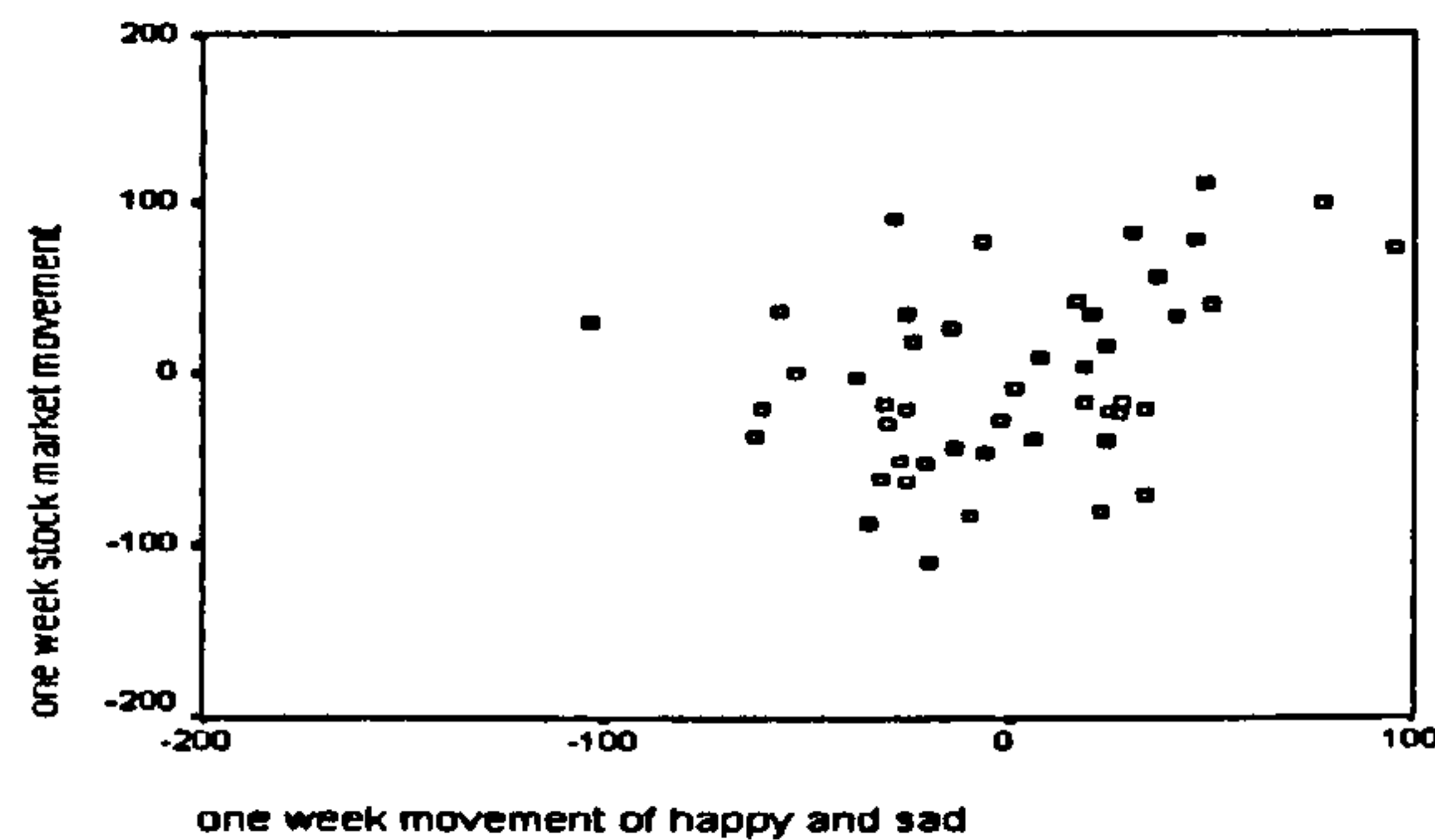


Figure 48: FTSE-ALL index movement and emotion movement

- **FTSE 100 data**

The values of diff_{ew} and diff_{sw} for FTSE 100's index data were found to be significantly correlated. The correlation result is 0.298, which is significant ($p=0.05$). The results are shown in figure 49.

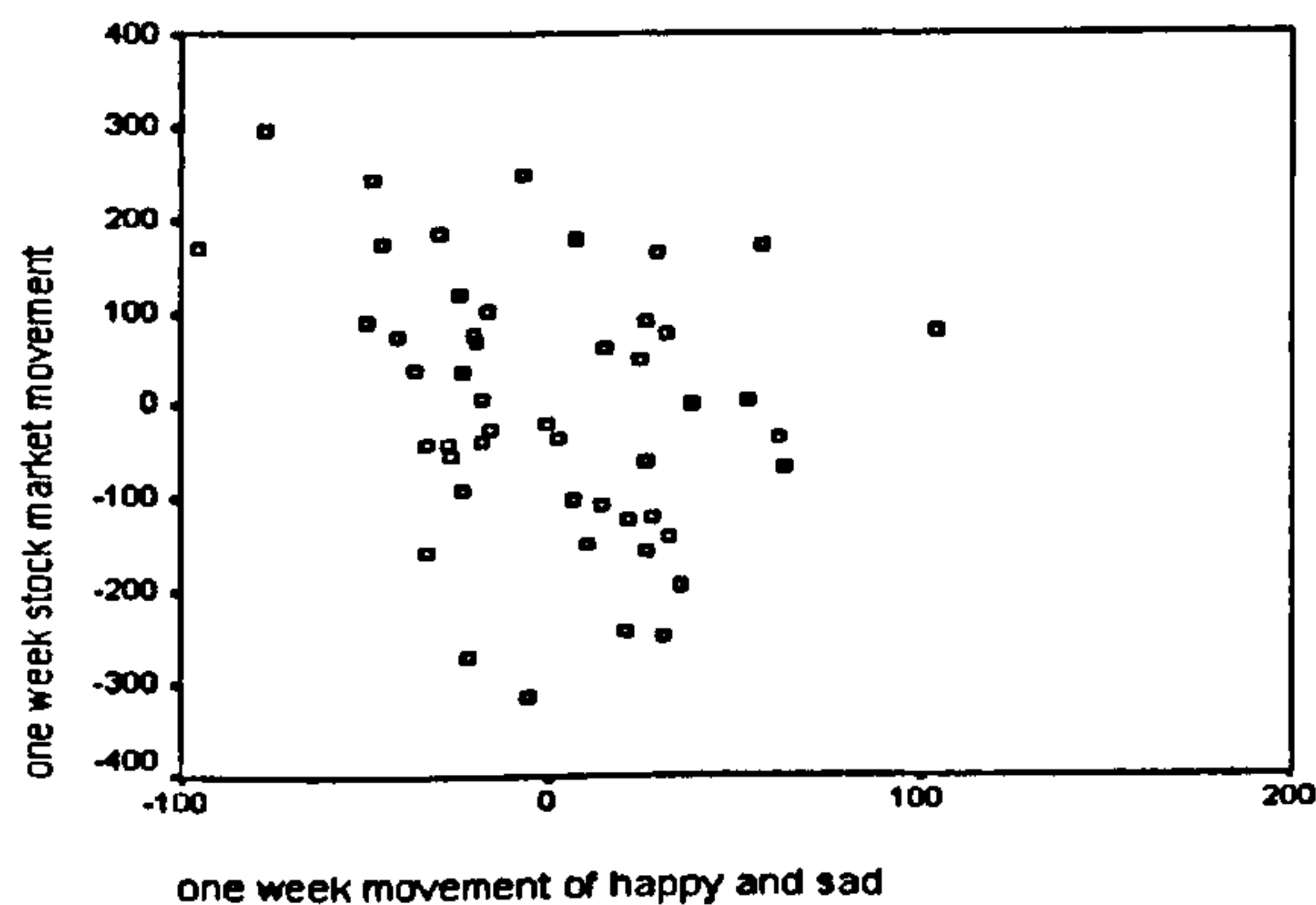


Figure 49: FTSE 100 movement and emotion movement

This test shows that in the weekly level, the movement of the intensity difference between happiness and sadness is significantly related to the stock market's next day's movement between the two weeks.

Test 3: divided data analysis

In this test, the emotion data was divided into two groups. The classification standard is the article's intensity difference between emotion happiness and sadness. If the absolute intensity difference is larger than 10, then the article is categorised as a large absolute difference group, otherwise it belongs to the small absolute difference group.

For the large absolute difference group, no significant correlation between the market movements and the intensity difference was found. For the small absolute difference group, a significant relationship between the next day's movements and the intensity difference was found in the FTSE-ALL index data. No significant correlation was found in FTSE 100 data. For the FTSE-ALL index test, the correlation result was 0.47, which was significant ($p=0.04$). The results are also shown in figure 50.

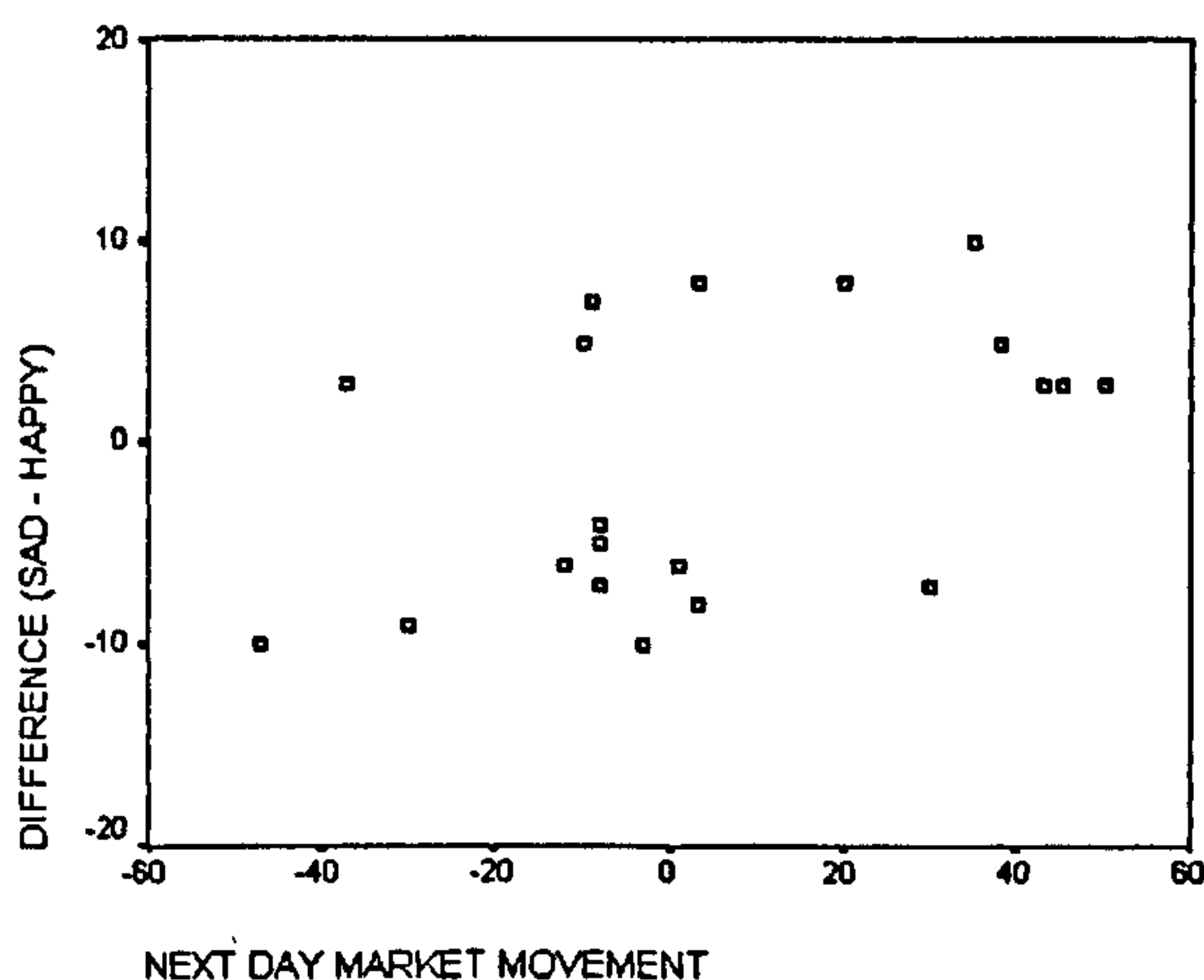


Figure 50: FTSE-all index result of small difference and the next day's market movement

This significant correlation found in FTSE-ALL index might mean that when the differences between the detected happiness and sadness emotions are small, the movement of market will be small. It is more likely that smaller movements in the stock market index will be sustained over longer periods of time and that these will be reflected in the emotional articles. For example, if the detected happiness-sadness=3, the next day's movement will possibly go up by a small amount. Since no significant correlation was found in FTSE 100, this relationship needs to be tested further.

For the large absolute difference group, no significant correlation was found. The reason may be that large movements in the market are relatively rare and less likely to occur in consecutive weeks. Large movements in the market may be reflected on the large differences between the happiness and sadness indicators in the articles; however this is unlikely to be accompanied by another large movement in the corresponding next day's stock market index.

7.1.6 Stock market emotion analyser conclusions

Emotional news aims to transform mathematical data into text and provide behaviour references to shareholders. A stock market emotion analyser has been developed successfully to analyse stock market articles and calculate the corresponding emotion status.

Experiments were carried out to test the relationships between the emotions detected by the stock market emotion analyser and share index value. The experiment results showed that more than 66% of articles correctly predicted the next day's movement trend. The experiment also shows that a significant relationship between the detected emotions and next day's overall value exists in the FTSE 100 and FTSE-ALL index level. When the absolute intensity difference between emotions happiness and sadness is large in an article, the next day's index value change will be large. Another finding is that the movement of the difference between happiness and sadness emotions in two articles in two consecutive weeks is significantly related to the stock market's next day's movement between the two weeks.

From the experiment results, it can be seen that emotions embedded in stock market articles do have significant influences on the stock market movements. A large number of articles reflect the way the stock market moves. These articles can be used to predict the next day's movement. Where there is a large difference between the intensities of happy and sad emotions presented in the articles, there will be a large movement in the market value in the next day, while where there is a small difference between the emotions there will be a smaller movement in the market value in the next day. However, only articles that contain a small difference between emotions may be used to predict the direction of the movement of the market value.

7.2 Chat environment applications

Currently, emotions in most chat interfaces are visualised by the manually selected emotion icons. The simplicity of emotion icons limits the expressions that can be presented, and as every user shares the same set of emotion icons, the interface can not be customised. By embedding

the emotion extraction engine into a chat interface, an expressive chat interface can be created. The 'Expressive chat interface' is a visual interface for effective real-time collaboration over the Internet and among groups of people. The 'Expressive chat interface' application has two variations: an 'Expressive chat 2D interface' and an 'Expressive chat 3D maze'.

By embedding the emotion extraction engine, the 'Expressive chat interfaces' are capable of invoking facial expressions without making use of video. The interfaces utilise discrete images of the participants in order to keep the bandwidth requirements to a minimum yet still provide an elaborate communication tool. The interfaces allow the viewing at a glance of participants in the system, and those pairs engaged in conversation, as well as the expressive image of the users engaged in the conversation. Three components: text, voice (synthesised by Microsoft speech engine) and expressive images are used in the interfaces. The expressive 2D interface and the expressive 3D interface are presented in figure 51 and 52.

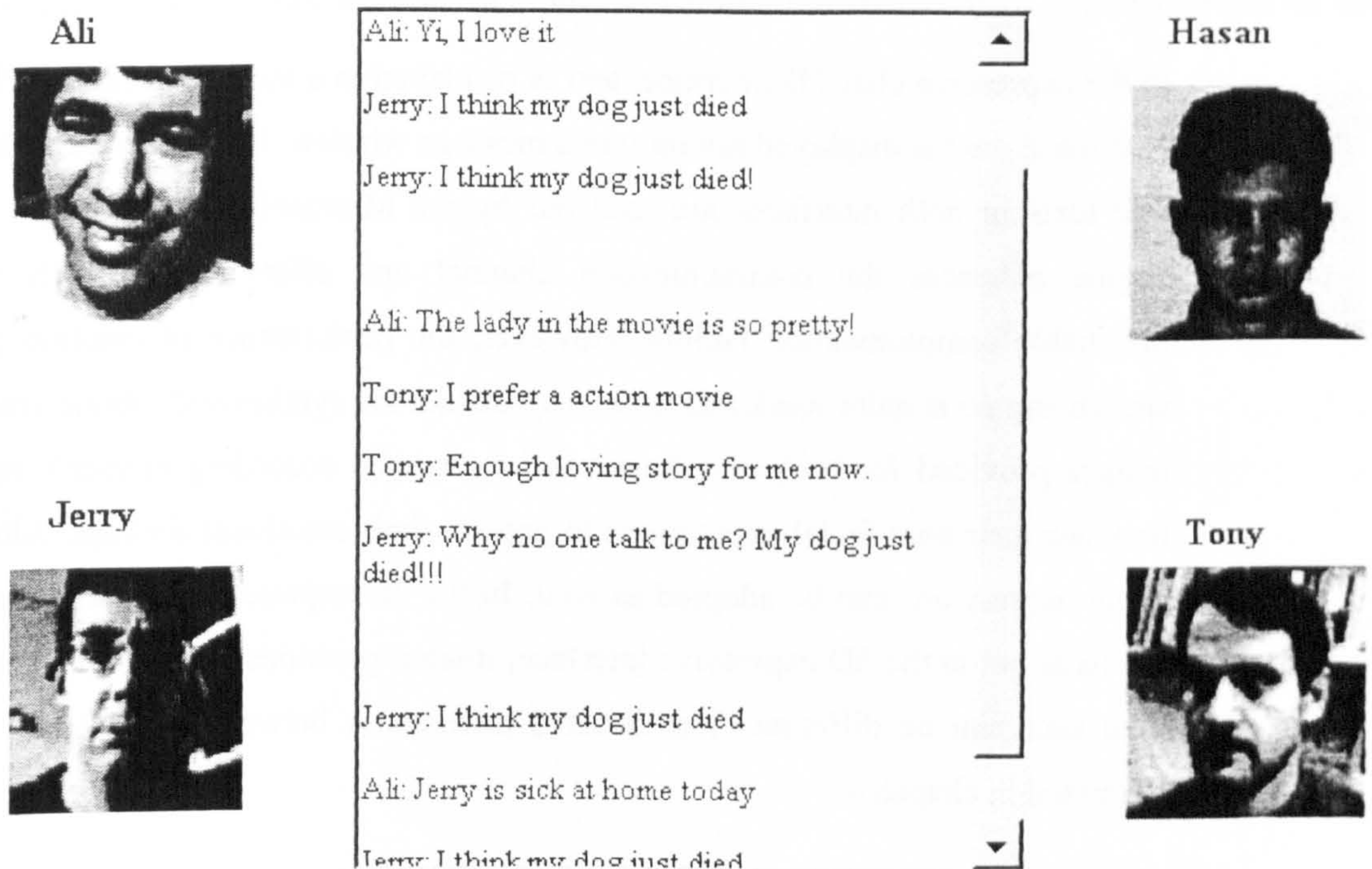


Figure 51: The expressive chat 2D interface

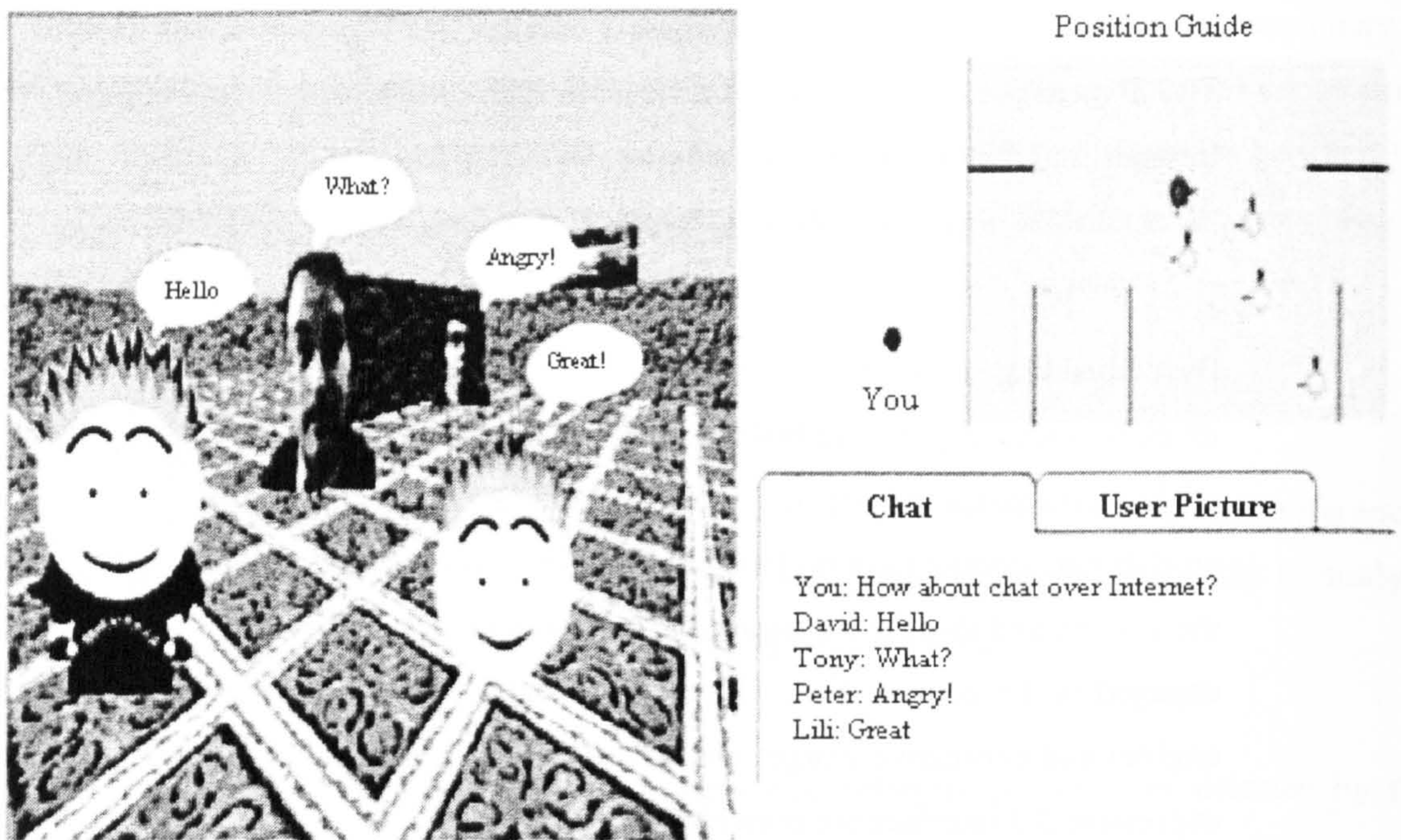


Figure 52: The expressive chat 3D maze interface

In the expressive chat 2D interface, text is displayed in a message window. However, in the 3D chat maze, text is displayed not only in a message window, but also in a text bubble in the maze. The texts in both interfaces are read out by the Microsoft speech engine. Using the speech engine enhances the communication channel and offers people with visual impairment invaluable communication facility. However, the performance of emotion presentation in the speech engine is quite weak and users may dislike the synthesised robotic voice. The expressive images provided for both interfaces can be changed according to user's requirements. Users may use their own facial expressions to present their emotional feelings. Alternatively, cartoon facial expressions can be adopted as well. In the 2D expressive interface, the images positions are fixed but in the 3D expressive interface, images positions are dynamic and the view point of each user can be different. The detailed differences between the 2D and 3D interfaces are discussed in chapter 9.

7.3 Chapter conclusion

This chapter describes different applications of the emotion extraction engine. A stock market emotion analyser, a 2D expressive chat interface and a 3D expressive chat interface are presented. The 2D expressive chat interface and the 3D expressive chat interface directly use the emotion extraction engine and the only difference is the interface layer. The stock market emotion analyser demonstrates how to revise the fundamental components, i.e., tagset and

tagged dictionary, to fit into the stock market context. By applying similar procedures, the emotion extraction engine may be able to be used in various different environments.

The stock market emotion market analyser itself revealed some interesting phenomena. By inputting stock market articles into the stock market emotion analyser, the emotional feeling of the article can be detected and the links between the emotions of the articles and the next day's stock market movements are clarified. The stock market emotion analyser may become a handy tool for non-expert stock market dealers.

The next chapter focuses on an experiment that assessed the preference of interfaces embedding the emotion extraction engine. The preference of different interface components (expressive images, voice and text) and component combinations (expressive image plus text, voice plus text etc.) were compared to detect whether users prefer an interface embedding the emotion extraction engine. Experiment participants are classified into groups according to their cognitive styles and statistical tests were carried out to examine the influence of cognitive style on the preference.

Chapter 8

Preference experiment on emotion extraction engine

In this chapter, an experiment assessing the preference of an interface embedding emotion extraction engine is presented. First the background knowledge related to the experiment (e.g. cognitive style etc.) is discussed and then the experiment platform is presented. The experiment results are illustrated in detail at the final part of this chapter.

8.1 Cognitive style

In this section, definitions of cognitive style (CS) are explored. The classification of CS used in the Cognitive Styles Analysis (CSA) program [Riding 1991 and Riding 1998] was chosen as the definition of CS used in the experiments for this research work.

8.1.1 Definitions of cognitive style

CS is the consistent underlying method of an individual's thinking and perceiving that subsequently affects the way in they perceive and respond to events and ideas [Tennant 1988, Riding 1991 and Riding 1998]. There is a wide range of different labels and methods of measuring CS. Riding and Cheema [1991] reviewed over 30 methods of defining CS and concluded most of the methods could be grouped within two fundamental independent CS dimensions – Wholist-Analytic dimension and Verbal-Imagery dimension.

8.1.2 The Wholist-Analytic and Verbal-Imagery dimensions of cognitive style

A person's position along the Wholist-Analytic dimension reflects whether they understand situations as a whole or see things in parts, while their position along the Verbal-Imagery dimension reflects the manner in which they represent information while thinking, either as words or mental pictures or images [Riding 1991 and Riding 1998]. The Wholist-Analytic and Verbal-Imagery groupings of style are represented as the independent dimensions shown in figure 54.

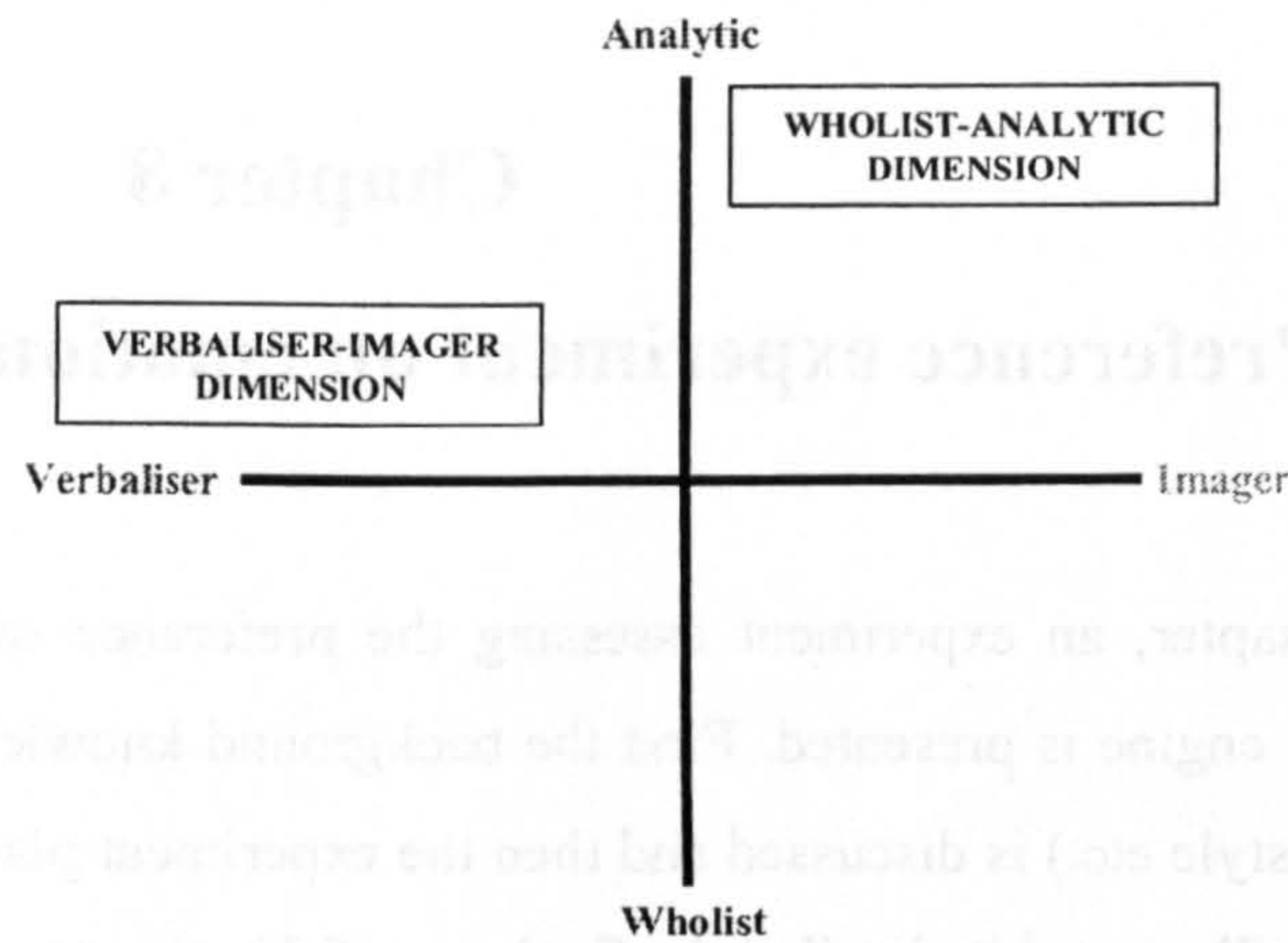


Figure 53: The two cognitive style dimensions

The characteristics of each CS (Wholist, Analytic, Verbaliser and Imager) determine the way individuals perform in social situations (whether they are outgoing or more restrained), at work situations (whether they are self-organised or prefer to be organised by others) as well as how individuals perceive and respond to information.

The Wholist-Analyst dimension determines whether people understand information or situations as a whole, or see things in parts. Wholists tend to view a situation as a whole and have an overall perspective. New information is stored serially in the brain as loosely clustered wholes rather than associated with related information that is already known. Analytics will tend to view a situation as a collection of parts, and will often focus on one or two aspects at a time to the exclusion of the others. Analytics separate out new information into conceptual groupings and store it in relation to what they already know.

The Verbaliser-Imager dimension determines the characteristic mode in which individuals represent information while thinking and the way they relate to social groups. Verbalisers consider the information they see, read or listen to, primarily as words or verbal associations (Verbaliser in figure 54). As speech is the basic medium of communicating with others, Verbalisers have an in-built advantage in social situations compared to Imagers and tend to be more out-going and socially aware.

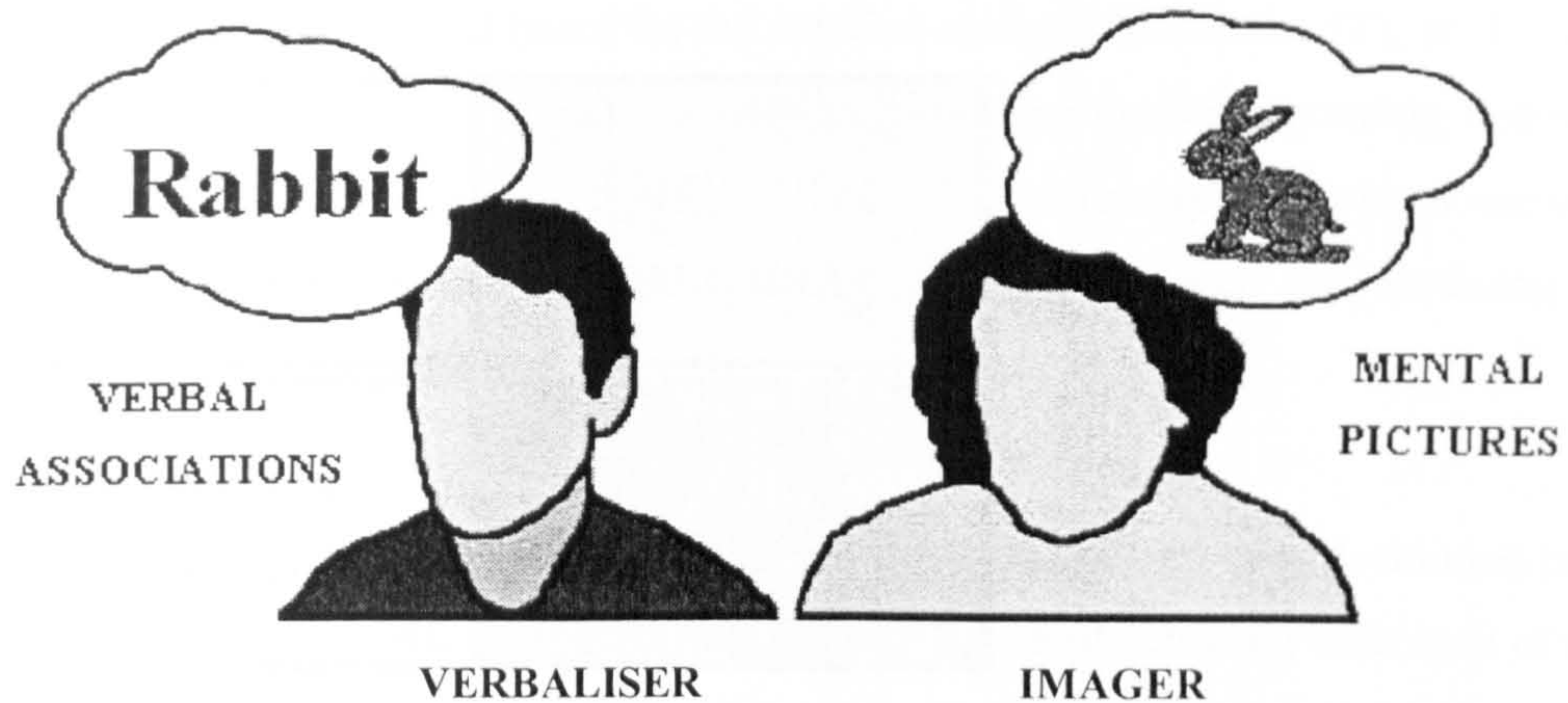


Figure 54: Verbaliser and Imager mode of representing information

8.1.3 The Cognitive Styles Analysis program

The Cognitive Styles Analysis (CSA) program was produced by Birmingham Learning and Training Technology [Riding 1991 and Riding 1998] in order to automatically calculate the CS of individuals in terms of their Wholist-Analytic and Verbal-Imagery dimensions. A person's position along each dimension is assessed by their performance when participating in a series of simple questions, half of which are designed to suit the characteristics of one style and half are designed to suit the opposite style.

The questions that assess the classification of the Wholist-Analytic dimension use graphical images. The Wholist style questions ask the participants to identify differences in the overall appearance of objects while the Analytic style questions ask participants to identify whether simple objects were contained within other more complex objects. The questions that assess the Verbal-Imagery classification of participants use simple statements comparing pairs of words. Participants are asked to determine whether they were *true* or *false*. Verbaliser style questions use low imaging words such as the names of concepts, while the Imager style questions use high-imagery words such as the names of objects that can be easily visualised [Riding 1991 and Riding 1998]. In experiments this classification is often simplified to produce four CS groups as shown in figure 55 [Riding and Sadler-Smith 1992].



Figure 55: Cognitive style quadrants

8.2 Planning the experiment

In conventional online communication (e.g., a chat room environment or an email application), text and voice are the major communication channels. With the emotion extraction engine, expressive images are introduced to the online communication as a new component and an expressive interface can be created. Individual's personality (e.g. cognitive styles) may influence their preferences of interfaces. The development of the emotion extraction engine leads to a number of questions that need to be answered, including:

- Do users prefer an interface with automatic expressive images display?
- Is the expressive image a helpful component in the interface design?
- Does the individual's cognitive style (CS) influence the preference of an interface with expressive image display?

8.3 Method of the experiments

The main aim of the experiment was to determine whether users prefer an interface embedding the emotion extraction engine to automatically display expressive images. To assess the preference, the expressive 2D chat interface described in chapter 7 was used as the test interface. The interface provides three components: text (T), voice (V) and expressive images (E). Other communication methods (e.g. video conference) exist in online communication environment, however, the expressive 2D interface does not use these methods as they are not widely used in online chat environments.

For the three different components - expressive image, voice and text (E, V and T), in total seven different component combinations exist: E, V, T, E+V+T, E+V, E+T and V+T. The

expressive image (E) is displayed based on the emotion analysis of the text (T), so $T \rightarrow E$. It is meaningless to show the expressive pictures without displaying the corresponding sentences, so combinations E and E+V are omitted. It is not common for Internet applications to use voice (V) as the only method to communicate, thus the combinations chosen for assessing in this experiment are E+V+T, E+T, V+T and T.

To compare the four combinations, four corresponding test interfaces were developed based on the expressive 2D interface. All the interfaces simulate two people chatting with each other. The expressive image (E) was generated by the expressive image generator, voice (V) was synthesised by Microsoft speech engine version 5. Instead of letting participants input sentences arbitrarily, the contents of the dialogues (T) were pre-designed. The reason was to control the emotions perceived by the participants and assure that each participant was presented with the same content.

To analyse whether participants' cognitive style (CS) influence the preference of the interface, participants were classified into groups according to their cognitive styles. The participant's CS was the between-participants independent variable and interface style was the within-participants independent variable. Totally there were four conditions depending on the CS styles of participants and these four conditions are shown in figure 56.

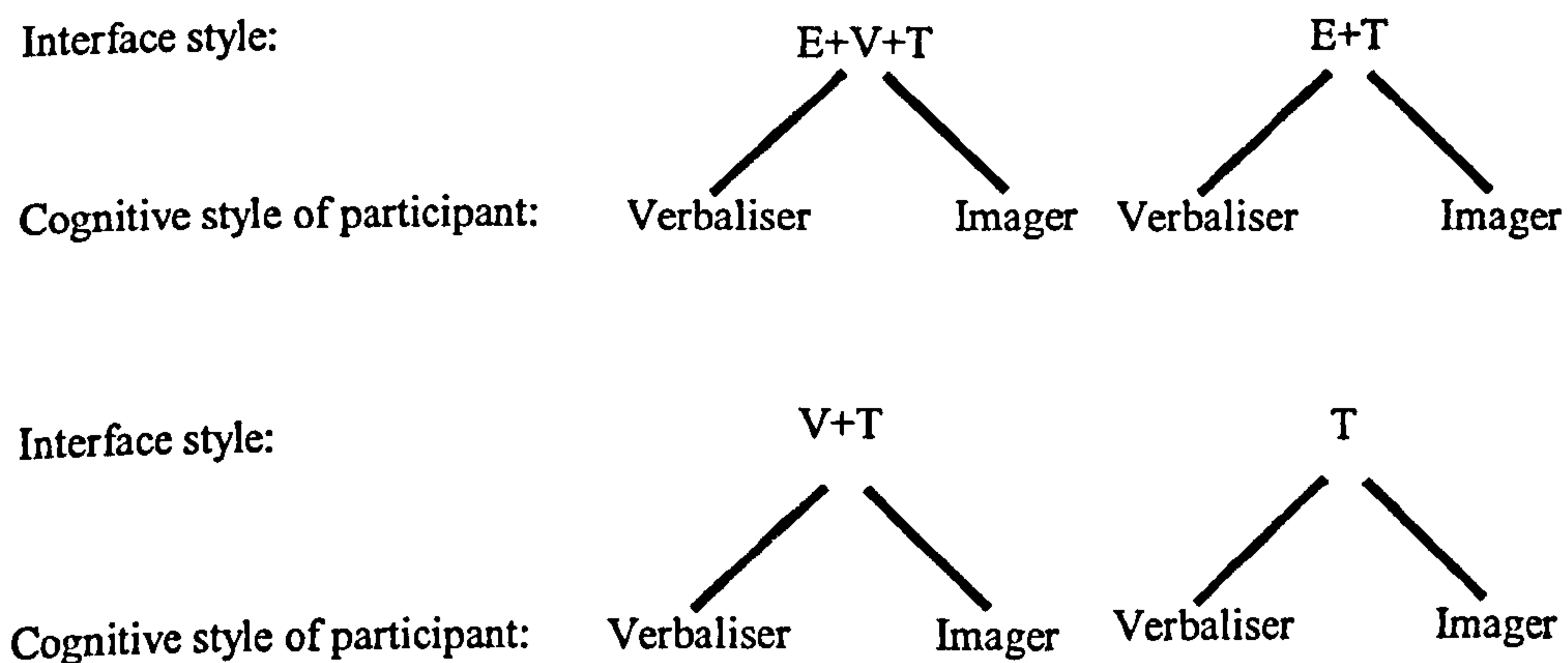
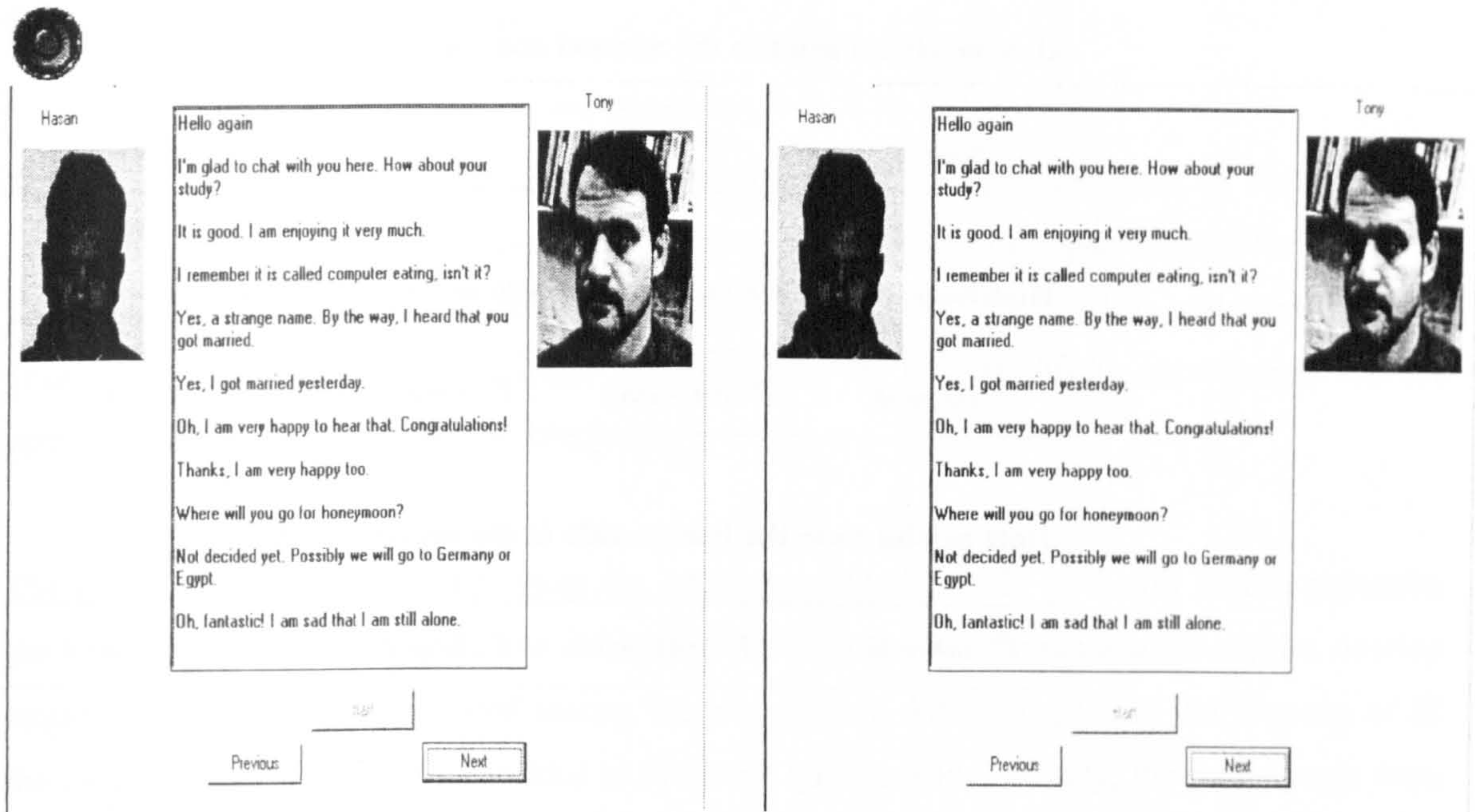


Figure 56: The mixed factorial fully crossed 2x4 design

The preference of the expressive interface was examined as a whole first and then the preference differences of different interface styles were examined separately for each CS dimension. Participants were asked to view all the four interfaces. Any effects that may have resulted because of the order which participants undertook were counter-balanced by

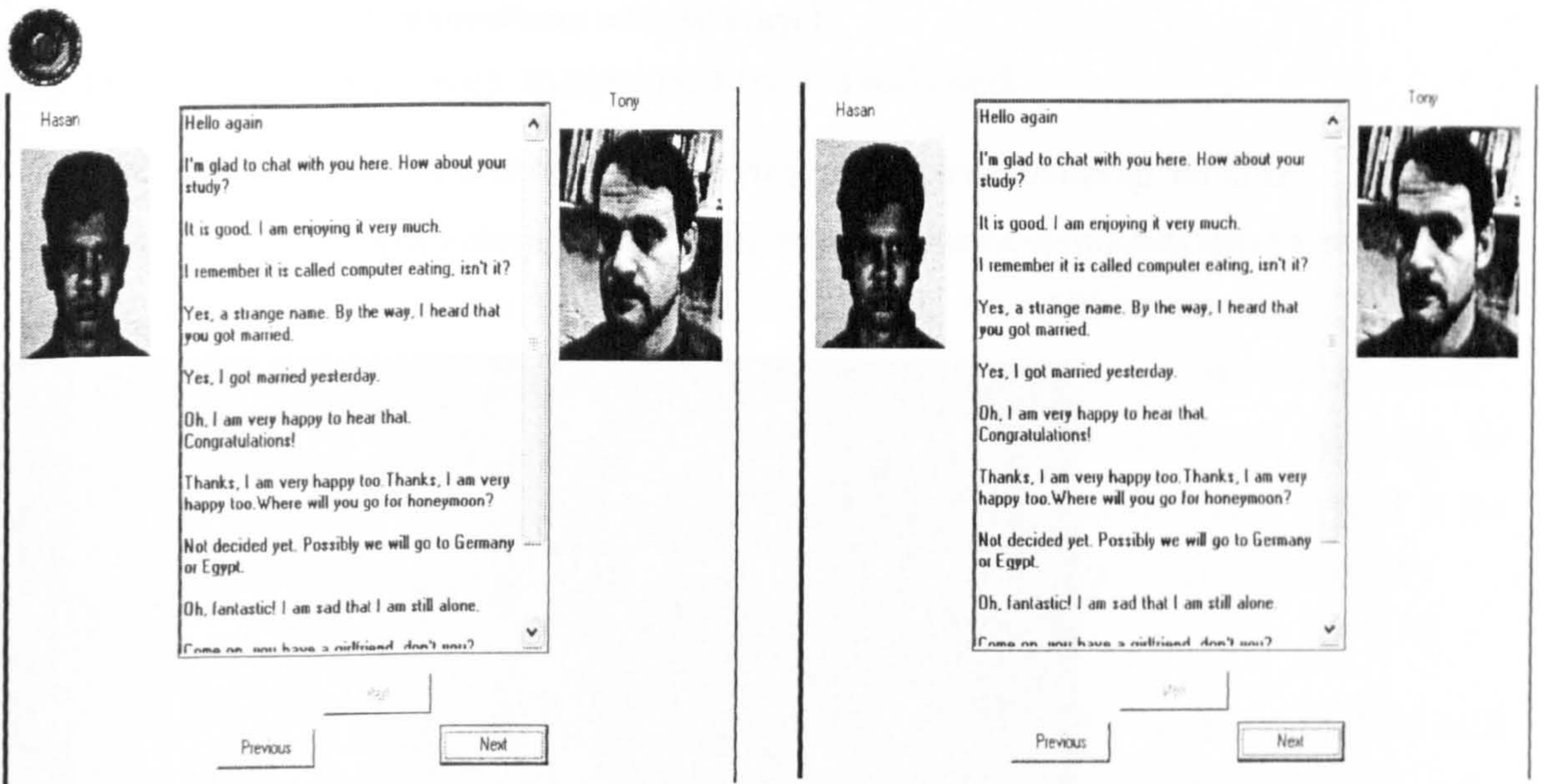
randomising the order (i.e. effects that participants viewed interface E+T+V first were counter-balanced by the other participants view interface V first).

All interfaces were implemented with the same screen layout as displayed in figure 57. The four interfaces contained the same amount of text, so the same amount of emotions was presented in each interface. The visual interfaces (font, size, button and layout, etc.) were identical across the four different presentations. The E+T+V interface contained three components: expressive images, voice and text, which mean that users could hear the voice, read the text and view the expressive images generated by the engine. The E+T interface did not include the voice component, so users could not hear the speech. The V+T interface consisted of voice and text thus users could not view the expressive images. The T interface only contained the text component and in consequence users could only read the text.



(a) Expressive image + voice + text

(b) Expressive image + text (no voice)



(c) Voice + text only (no expressive image)

(d) Text only (no expressive image and voice)

Figure 57: The four media combination interfaces

After viewing these four interfaces, participants were redirected to an online questionnaire. The questionnaire contained questions assessing participants' preferences of every component and each interface, e.g., do you like the E+T+V interface? In figure 58, a typical screen of the online questionnaire session is shown.

How much did you like the second application (emotion+text)

liked very much like dislike disliked very much

How helpful were the emotions provide in the second application?

not at all not helpful helpful very helpful

How helpful were the text provide in the second application?

not at all not helpful helpful very helpful

Figure 58: The questionnaire session

After the questionnaire session, participants were asked to carry out the CSA test. A snap shot of the CSA test is shown in figure 59.

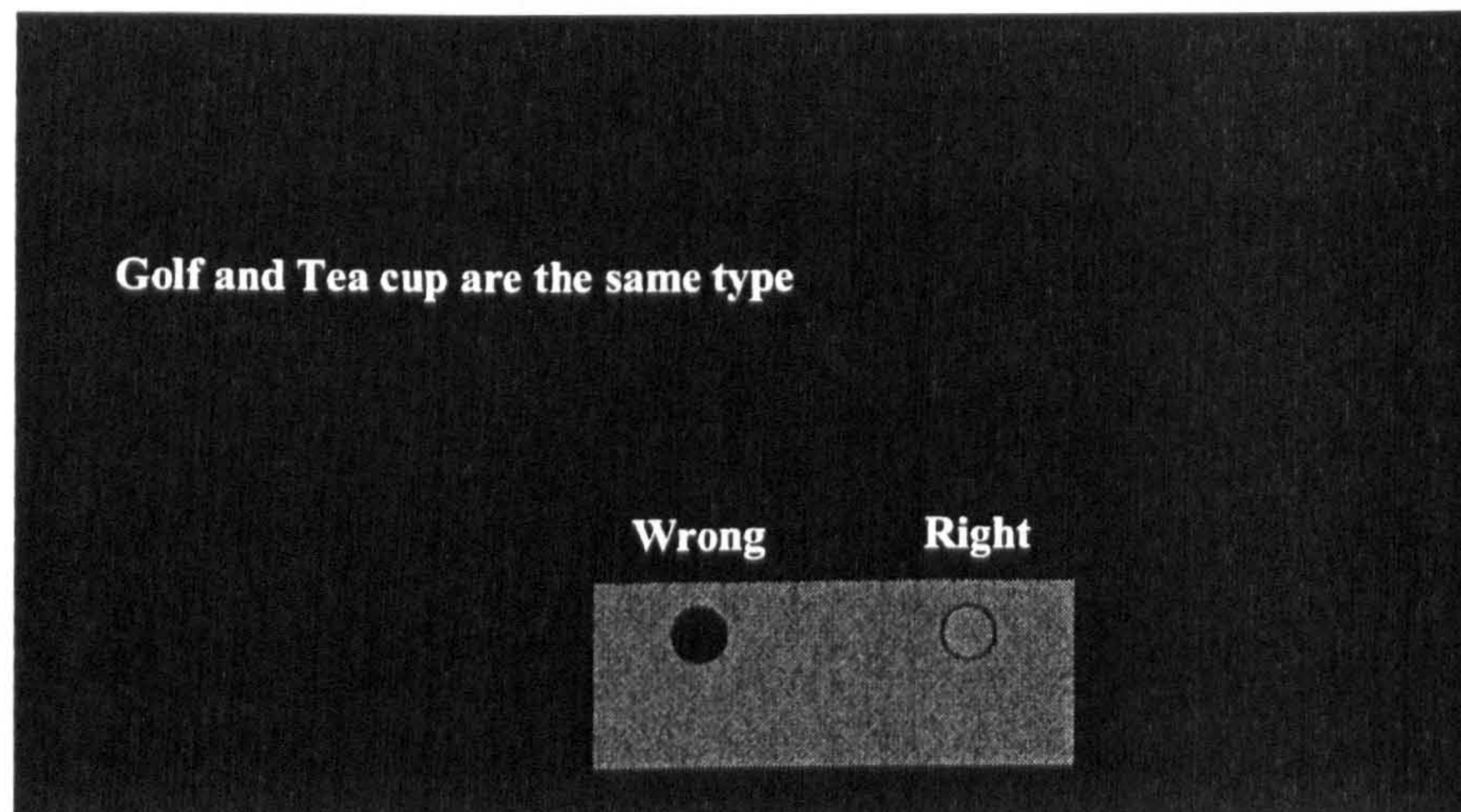


Figure 59: The CSA session

After answering all the questions in the CSA session, the CSA software will automatically determine individual participant's CS. The output format is a pair of values, one indicates their Wholist-Analytic classification, and the other indicates their Verbal-Imagery classification.

8.4 Experiment operation

A total of 50 participants were selected at random from students and staff at Bournemouth University. When starting the experiment, the experimenter told each participant that she or he will view four interfaces that simulate two people chatting with each other, and then fill in an online questionnaire. The experimenter also explained the purpose of the experiments and the operational procedures for the test interfaces.

The participants were directed to view the interfaces E+V+T, E+T, V+T and T. To randomise the viewing sequence, an application was developed to ensure the randomisation, so the viewing sequences were evenly distributed among the participants. After completing the viewing of all the interfaces, participants were guided to fill in the questionnaire. Finally, the participants were asked to run the CSA package to determine their CS.

8.5 How preference was measured and analysed

There were two main measurements of the preference comparisons in this experiment. The first was the amount of preference that participants showed to different interfaces. This was measured by analysing the participants' responses to the question "Do you like this interface?". The second measurement was the preference of each component. This was measured by analysing the participants responses to the questions "Are the expressive images helpful in the interface?", "Is the voice helpful in the interface?" and "Is the text helpful in the interface?".

Four selections - "not at all", "not helpful", "helpful" and "very helpful" were provided for each question and values -2, -1, 1 and 2 were assigned correspondingly. When calculating the percentage of the preference, "not at all" and "not helpful" would be classified as "not helpful", in contrast, "helpful" and "very helpful" would be categorised as "helpful".

To test whether an individual's CS influences the preference of different interface and individual components, participants were divided into groups according to their CS and a between group comparison was carried out to determine whether CS played an important role in the preference of different interfaces and different components. It was hypothesised that CS may strongly influence the preference of different interfaces and components, and individuals whose CS belongs to Imager category would prefer interfaces with expressive images more. Each CS

group was compared against each other using a number of procedures to test the statistical significance of the observed differences.

8.6 Experiment results

This section presents the collected feedback from the questionnaire and examines whether participants prefer the interface with expressive images and whether the preference of the participants support the hypothesis.

8.6.1 Question analysis

Question 1: Preference of interfaces

For the question: “Do you like the interface (E+V+T)?”, 39 participants answered “like” and 11 participants’ answered “dislike”. For the question: “Do you like the interface application (E+T)?”, 38 participants’ chose the option “like” and 12 participants chose the option “dislike”. For the question: “Do you like the interface (V+T)?” 34 participants answered “like” and 16 participants answered “dislike”. For the question: “Do you like the interface (T)?” 25 participants chose the answer “like” and 25 participants chose the option “dislike”. The preferences of different interfaces are shown in table 26. It can be shown that the interface (E+V+T) is the most preferred.

	Like (E+V+T)	Like (E+T)	Like (V+T)	Like (T)
Participants number	39	38	34	25
Participant percentage	78%	76%	68%	50%

Table 26: Preferences of different interfaces

Conclusions cannot be drawn from the percentage analysis without testing whether the preference of each interface was significantly different. Two-way analysis of variance (ANOVA) calculations were carried out on the above results using SPSS. The statistic analysis results are shown in table 27. The columns show the T statistic, the level of significance (sig) and degrees of freedom (df).

Tasks	T	Df	Sig
Like (E+V+T) vs. like (T)	2.307	49	0.025
Like (E+V+T) vs. like (E+T)	0.781	49	0.438
Like (E+V+T) vs. like (V+T)	1.155	49	0.254
Like (E+T) vs. like (T)	1.949	49	0.058
Like (E+T) vs. like (V+T)	0.425	49	0.673
Like (V+T) vs. like (T)	1.606	49	0.115

Table 27: 2-Way ANOVA results of interface preferences

It can be seen that the significant difference exists between the preference of interface (E+V+T) and the preference of interface (T) ($p=0.025$). A marginal significant difference exist between the preference of interface (E+T) and the preference of interface (T) ($p=0.058$).

It was expected that the interface with all components (expressive image, voice and text) would be the most preferred interface. This hypothesis was partially supported by the preference analysis and the significant preference difference between the interface (E+V+T) and interface (T). The significant preference difference and the marginally significant preference difference also demonstrate the user satisfaction with component E (expressive image).

Question 2: Helpfulness of expressive images

For the question “Are the expressive images helpful in the interface (E+V+T)”, 36 participants selected the answer “helpful”. For question “Are the expressive images helpful in the interface (E+T)”, 35 participants selected answer “helpful”. The preference analysis showed that most participants agree that the expressive images are helpful in both interfaces. The results are shown in table 28.

	Helpfulness of E in E+V+T	Helpfulness of E in E+T
Participant number	36	38
Participant percentage	72%	76%

Table 28: Helpfulness of expressive images in the E+V+T and E+T interfaces

A sceptical view of the helpfulness of a component can be summed up as follows: half think helpful and half think unhelpful. For the above analysis, is this outcome is significantly different

from that which is expected under half think helpful and half think unhelpful? Chi-square tests were carried out and results are shown in table 29 and 30. The columns show the chi-square statistic, the degrees of freedom (df) and the level of significance (sig). It can be seen that the in interface E+V+T, a significant difference exist between the half and half assumption ($p=0.002$). In interface (E+T), a significant difference exists ($p=0.000$) as well.

	Helpfulness of E in E+V+T
Chi-Square	9.680
Df	1
Sig	.002

Table 29: Chi-Square results for the helpfulness of expressive images in the E+V+T interface

	Helpfulness of E in E+T
Chi-Square	13.520
Df	1
Sig	.000

Table 30: Chi-Square results for the helpfulness of expressive images in the E+T interface

Question 3: Helpfulness of voice

For the questions “Is the voice helpful in the interface E+V+T”, and “Is the voice helpful in the application V+T”, the preference analysis show very similar results. The percentage values of the voice preference in different test interfaces are shown in table 31.

	Helpfulness of voice in E+V+T	Helpfulness of voice in V+T
Participant number	35	38
Participant percentage	70%	76%

Table 31: Helpfulness of voice in the E+V+T and V+T interfaces

The percentage analysis showed that 70% of participants believed that voice is helpful in interface E+T+V and 75% participants agree that voice is helpful in interface V+T. Chi-square tests were carried out and the results are shown in table 32 and 33. It can be seen that the in interface E+V+T, significant differences exist between the half and half assumption ($p=0.003$ and $p=0.000$ correspondingly).

	Helpfulness of V in E+V+T
Chi-Square	8.320
Df	1
Sig	.003

Table 32: Chi-Square results for the helpfulness of voice in the E+V+T interface

	Helpfulness of V in V+T
Chi-Square	16.43
Df	1
Sig	.000

Table 33: Chi-Square results for the helpfulness of voice in the V+T interface

Question 4: Helpfulness of text

The percentages of participants choosing text as helpful in the different interfaces are very similar (all around 72%). However the mean values of the different interfaces are different. The mean values of the helpfulness of text among the different interfaces are shown in figure 60. It can be seen that text is more preferred in the interface E+T than other two interfaces E+V+T and V+T. To detect the possible significant differences between the mean ratings, 2-way ANOVA analysis was carried out. significant differences of users perceptions of the helpfulness of text are found between interface E+T and interface T+V ($p=0.011$), which is shown in table 34.

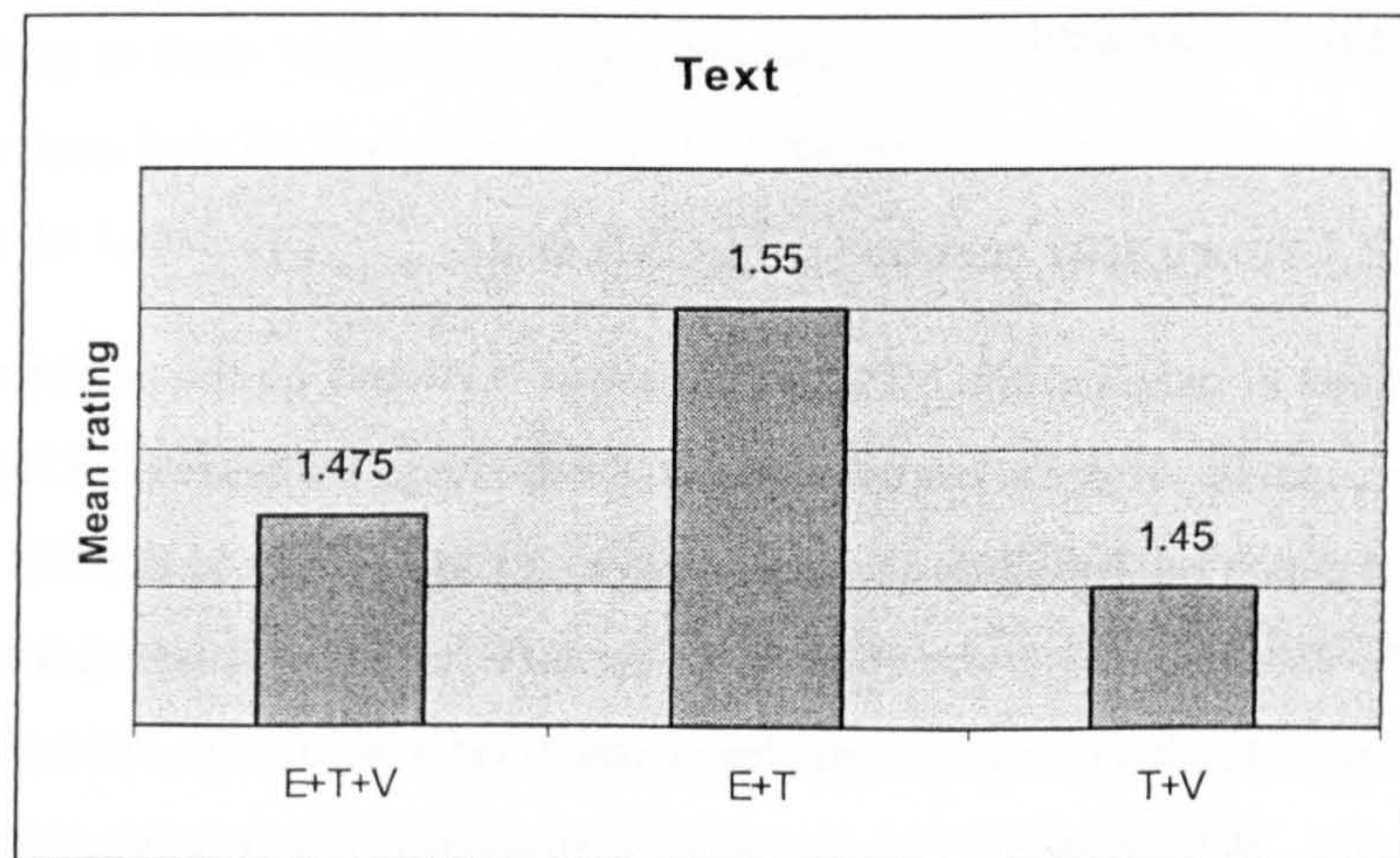


Figure 60: Helpfulness of text in the E+V+T, E+T and T+V interfaces

Tasks	T	Df	Sig
Helpfulness of T in (E+T) vs. Helpfulness of T in (T+V)	2.431	49	0.011

Table 34: Helpfulness of text 2-way ANOVA results

Question 5: Component preferences across applications

For the question “In all the four interfaces, which component do you prefer most?”, 17 participants selected component “expressive image”, 20 participants select answer “text” and 13 participants chose “voice”. The results are shown in figure 61.

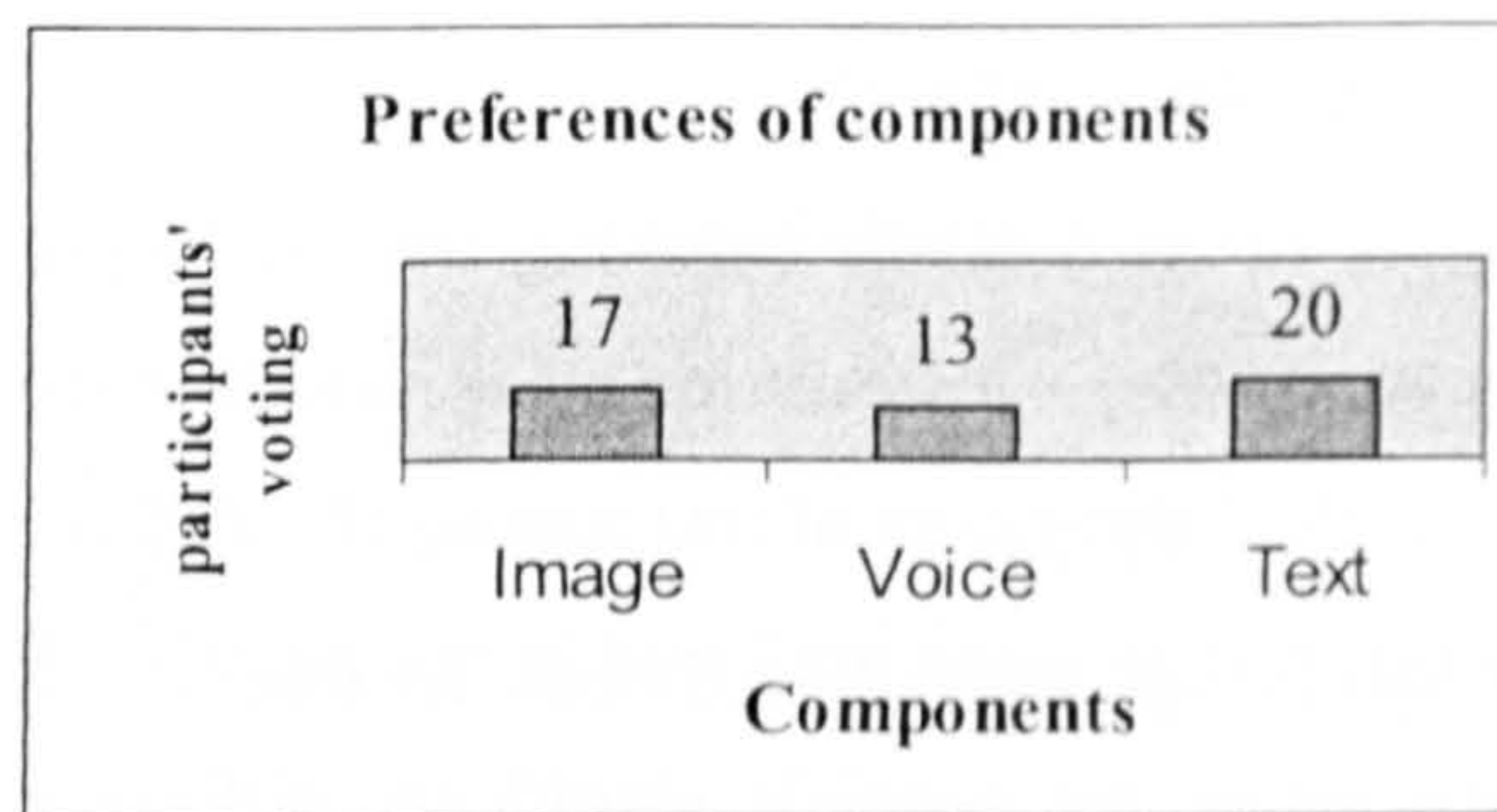


Figure 61: Preferences of the components

No significant difference was found after performing 2-way ANOVA analysis, but from the result, it may be concluded that “many participants choose the expressive image as the most preferred component”.

8.6.2 Experiment results from CSA tests

In order to examine the effect of cognitive style has on the performance of the participants, the mean results of each cognitive style group were compared. Dividing participants at the mid-point of the Intermediate section of the WA dimension and the mid point of the Bimodal section of the VI created four groups consisting of Wholists, Analytics, Verbalisers and Imagers. When the two dimensions are considered together a grid with four cognitive style quadrants are created; Wholist/Verbalisers, Wholist/Imagers, Analytic/Verbalisers and Analytic/Imagers. Experiment participants were plotted by their CSA rating for both dimensions in figure 62.

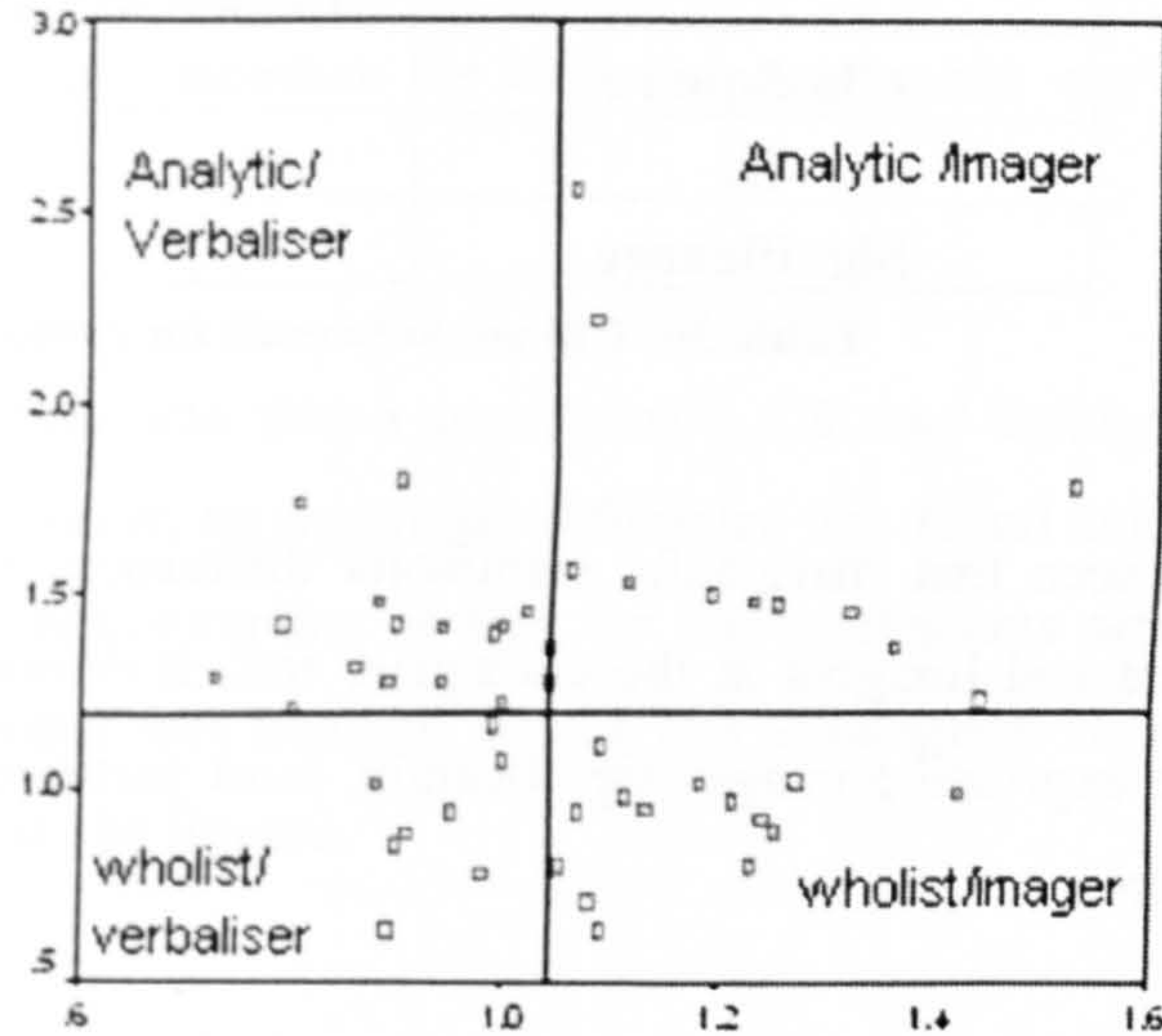


Figure 62: Scatter-graph of CS

As shown in figure 65, there were similar numbers of Wholists (22) and Analytics (28), and almost an equal numbers of Verbalisers (26) and Imagers (24). Dividing the participants into the quadrants there were 8 Wholistic/Verbalisers, 14 Wholistic/Imagers, 17 Analytic/Verbalisers and 11 Analytic/Imagers.

8.6.2.1 Interface preferences between different cognitive style groups

There are no significant differences between the different Wholist and Analytic groups in the interface preferences, as most participants prefer the interface E+T+V. When participants are divided according to their Verbal-Imagery classification, a significant preference difference is found between interface E+T and interface T. Imagers significantly prefer the interface E+T ($p=0.028$) than the interface T.

8.6.2.2 Component preferences by cognitive style

No significant difference of component preferences was found in the between-group analysis, for those 17 participants who preferred images, it can be shown that 12 (70.6%) of participants were Imagers and only 5 (29.4%) were Verbalisers. Table 35 presents the percentage information and the results of Chi-square test are shown in the table 36.

	Imager group	Verbaliser group
Participant number	12	5
Participant percentage	70.6%	29.4%

Table 35: Image preference by CS

	Imager group vs. Verbaliser group
Chi-Square	2.882
Df	1
Significance	.090

Table 36: Chi-square result for component preference

It can be seen that marginally significant differences are found between the preferences of Verbalisers and Imagers in the chi-square test. It demonstrates that although not all Imagers prefer the expressive images significantly, most participants who prefer the image component are Imagers.

8.7 Chapter conclusions

An emotion extraction engine was developed and the performance was assessed in previous chapters. However do people prefer an interface embedding the emotion extraction engine? This chapter presents an experiment setting up to test the assumptions that interfaces containing expressive images are preferred by participants, expressive image is a helpful component in interface design and participants' CS style may affect the preferences of different interfaces.

It was expected that participants would prefer interfaces that embedded the emotion extraction engine to automatically display expressive images. It can be seen that the interface E+V+T was preferred more significantly than the interface T and there was a marginally significant preference for the interface E+T than the interface T. However no significant difference was found between the preferences of the interface E+V+T and the interface V+T or between the interface E+T and the interface V+T. The expectation was only partially supported.

Another assumption was that the use of expressive images forms a helpful component in interface design. An overwhelming number of participants (74%) preferred the use of expressive images and a significant difference between the preferences of different components was found in chi-square test results.

The first CS assumption was that a participants' CS may influence their preference of interfaces and preference of interface component. No significant differences were found between the different wholist and Analytic groups in the interface preference or component preference test. The Verbal-Imagery between group test results showed a significant difference in the interface

preference - interface E+T was preferred significantly more by Imagers than Verbalisers. However, similar test using interface E+V+T did not reveal any significant preference difference.

The second CS assumption was that a participant's CS may influence their preference of interface components. However, no significant difference was found in most measurements. The only evidence support this assumption is that for the participants who preferred expressive images, a significant number was Imagers. Hence this assumption is only partially supported and the influence of CS may be limited.

The test platform described in this chapter is a 2D expressive interface. Will the preference differ between a 2D and a 3D interface? The next chapter will focus on the preference difference between a 2D expressive interface and a 3D expressive interface. Social factors, such as activity style, interface style and sociability will be examined.

Chapter 9

An experiment on assessing the preference difference between 2D and 3D expressive interfaces

The experiment results presented in the previous chapter suggested that people prefer the interface with expressive image display. The test interface was a 2D expressive interface. As the characteristics between 2D interface and 3D interface are different, will the preference of users differ? This chapter describes an experiment that was set up to examine the preference difference between 2D and 3D expressive interfaces. Three important characteristics of human-computer interaction: activity style, interface style and sociability were used as experiment variables to assess the difference in preference.

In this chapter, the background considerations of the experiment, including the definitions of social presence, activity style and human sociability, are discussed at the beginning. Then the questions that the experiment was attempting to answer and the method that was undertaken in conducting the experiment are presented. The different characteristics between a 2D expressive interface and a 3D expressive interface are discussed in detail and the experiment results are discussed in final section of this chapter.

9.1 Background

In a very general level, an Internet communication interaction has three important factors:

1. **Place:** Where is the interaction happened (e.g. a chat room, an online game room or a company's internal web site)?
2. **People:** Who are involved in the interaction (e.g. college students or wealthy millionaires)?
3. **Purpose:** What are the purposes of the interaction (e.g. spending leisure time, solving an urgent problem or looking for new friend)?

The above three factors may be mapped into three variables: interface style, Human sociability, and activity style.

Graphically, Internet communication interfaces can be classified into two categories: two-dimensional (2D) and three-dimensional (3D). A 2D interface is an acceptable choice for the monitor. 3D interfaces apply various graphical algorithms to simulate the sense of depth on the

monitor; hence most 3D interfaces can be defined as 2.5D. In this chapter, the 3D interfaces mentioned below can actually be classified into 2.5D. 2D and 3D interfaces may present different emotions and may be suitable for different purposes. In this section, the definition of social presence, human style and activity style are discussed.

9.1.1 Social presence

Communication channels are vivid in face-to-face communication. Physical movement, facial expressions and variations of sound create the diversity. Computer and the Internet communications cannot provide the physical presence of users. However, in the Internet, people feel that they are communicating directly with other users and one example is the chatroom application. This feeling is called social presence [Xiong 2002].

Social presence is defined as “The degree of salience of the other person in the interaction and the consequent salience (and perceived intimacy and immediacy) of the interpersonal relationships” [Short et al. 1976]. Communication researchers [Short et al. 1976 and Bailenson et al. 2001] argue that even in a text-dominated environment, social presence still exists and provides important functions.

Interfaces with rich or limited communication channels may lead to different amounts of perceived social feelings. Witmer and Singer [1998] discussed fourteen factors influencing social presence. These factors include degree of control, environmental richness, multi-modal presentation, scene realism, immediacy of control, anticipation, mode of control, physical modifiability, sensory modality, degree of movement perception, active search, isolation, selective attention, interface awareness and meaningfulness of experience. With social presence theory, different interfaces can be classified and assessed by the amount of social feelings presented. Jensen and colleagues [Jensen et al. 2000] argued that increasing social presence can improve enjoyment for an activity. As 2D and 3D interfaces may present different communication channels to users, the perceived social presence feelings will be different.

9.1.2 Human sociability style

Sociability is defined as the quality or state of being sociable. The Marriam-Webster online dictionary defines "sociable" as the incline by nature to companionship with others of the same species. Sociability is an important factor that differentiates humans [Nye and Brower 1996]. The same events may trigger significantly different feelings and actions because of the individual's sociability differences.

Various views exist over human sociability and three most important opinions are discussed in general here. One view is that human sociability is a gift from nature and so sociability is a general and universal goal which individuals naturally aspire to attain; a second view is that the sociability is an obligation, which means that sociability is a casual and accidental phenomenon, a secondary and not a primary objective; The third view is the sociability can be chosen and is determined by the individual's faculty of reasoning and calculation [Murtadha 1975].

It is agreed in most views that an individual's sociability has an effect on the preference of life style and may influence their actions and scene preferences. Some may enjoy going out socialising with friends, while others prefer reading books alone. The differences in social preferences may further influence the choice of Internet communication interface and the preference of the quantity of social presence feelings. Sociability may be considered as a reality in a double sense. On one hand sociability directly stimulates the individuals in their instinct to choose the preferred environment. On the other hand, there are interests connected with the activities that motivate such a choice.

9.1.3 Activity style

It is argued in this research work that the purpose of communication can be classified into two general categories: business-oriented and social-oriented. For business-oriented communication, people intend to grasp the information they need as soon as possible. On other hand, people choose social-oriented communication to make friends, set up relationships and create social networks. Table 37 lists some typical business-oriented activities and social-oriented activities.

Business-oriented activities	Social-oriented activities
Technique training	Take a break from work
Schedule technical meetings	Birthday party
Seek technical advice	Gossip chat

Table 37: The different categories of activities

Following usability principles and task-analysis theories [Badre 2002, Cato 2001, Dix et al. 1998, McCracken and Wolfe 2004, Nielsen 2000, Nielsen and Tahir 2002 and Preece et al. 2002], it can be predicted that business-oriented systems and social-oriented systems will require different types of interface - business systems are concerned with the efficiency of performing tasks, while the effectiveness of social-oriented systems depend more on users satisfaction and the experience of interacting with the system.

9.2 Planning the experiment

The problem the experiment was attempting to solve was to discover the possible preference differences between a 2D expressive interface and a 3D expressive interface. In order to plan the experiment, following considerations need to be examined. First, social presence feelings are important factors of the Internet communication interface and it may vary in different interfaces. How many social presence feelings can be perceived in the 2D and 3D expressive interface? Second, both activity styles and the individual's sociability style influence interface preferences and have different requirements on social presence feelings. Which factor is more important? - The activity style or the sociability? It was hypothesised that when an individual had a particular aim, e.g., solving a crucial problem or finding friends to talk, the activity style would dominate the preference of interface. The following hypotheses were developed based on the above discussions and the experiment was set up to examine whether these 3 hypotheses were correct.

Hypothesis 1: Activity style will strongly influence the preference of the 2D and 3D expressive interfaces.

Hypothesis 2: The individual's sociability will strongly influence the overall preference of the 2D and 3D expressive interface.

Hypothesis 3: The individual's sociability style will not strongly influence the preference of individual interfaces when the activity is chosen; hence the activity style is the dominant factor for the interface preference.

9.3 Experiment platform development

The 2D expressive chat interface and the 3D expressive chat interface were adopted as the test platform. The 2D and 3D interfaces only apply commonly available techniques, i.e., textual messages, textual bubbles and expressive images, to satisfy the universality requirements for experiment assessment. Automatic facial expression display is achieved in both the 2D and the 3D interface by integrating the emotion extraction engine.

The 2D interface was designed as a relatively simple environment. Similar to current online chat rooms, the 2D interface relies on textual messages. However the user's expressive images can be automatically displayed. After logging onto the 2D interface, user can easily find other users

and get an overview of the whole environment. The layout of the 2D interface has been shown in figure 51 in chapter 7.

The design idea of the 3D interface emerged from real-time 3D games and the interface is converted from a 3D maze. The interface is divided into different spaces by walls and users can move around with the aid of cursors. The 3D interface displays the current relative positions of all users within a "position guide panel". To engage in conversation with others, a user needs to walk close enough to another user. When a conversation starts, textual messages will be shown both within a message box outside of the 3D space and in a dialog bubble within the environment. To represent each user, facial images are displayed and users can choose to use either cartoon images or human images to represent them. For a snap shot of the 3D interface, please refer to figure 52 in chapter 7.

9.4 Social presence differences between 2D and 3D interfaces

In this section, the possible different social presence feelings perceived by experiment participants in the 2D and 3D expressive interface are described. In total, eight different senses are discussed.

1. Movement senses

Similar to most chat environments, every user in the 2D-chat interface is in a fixed position. In contrast, the 3D interface provides some aspects of movement. Users not only can move around the space, but also have specific field of view and can look for the spatial guidelines from the position guide panel.

2. Geographic senses

Unlike the 2D chat interface, a 3D interface can present various complex geographic entities, e.g., a city, or simple entities, e.g., a tree. Users may perceive geographic movement phenomenon during the virtual movement. Figure 64 shows the field of view and the position guide component of the 3D interface.

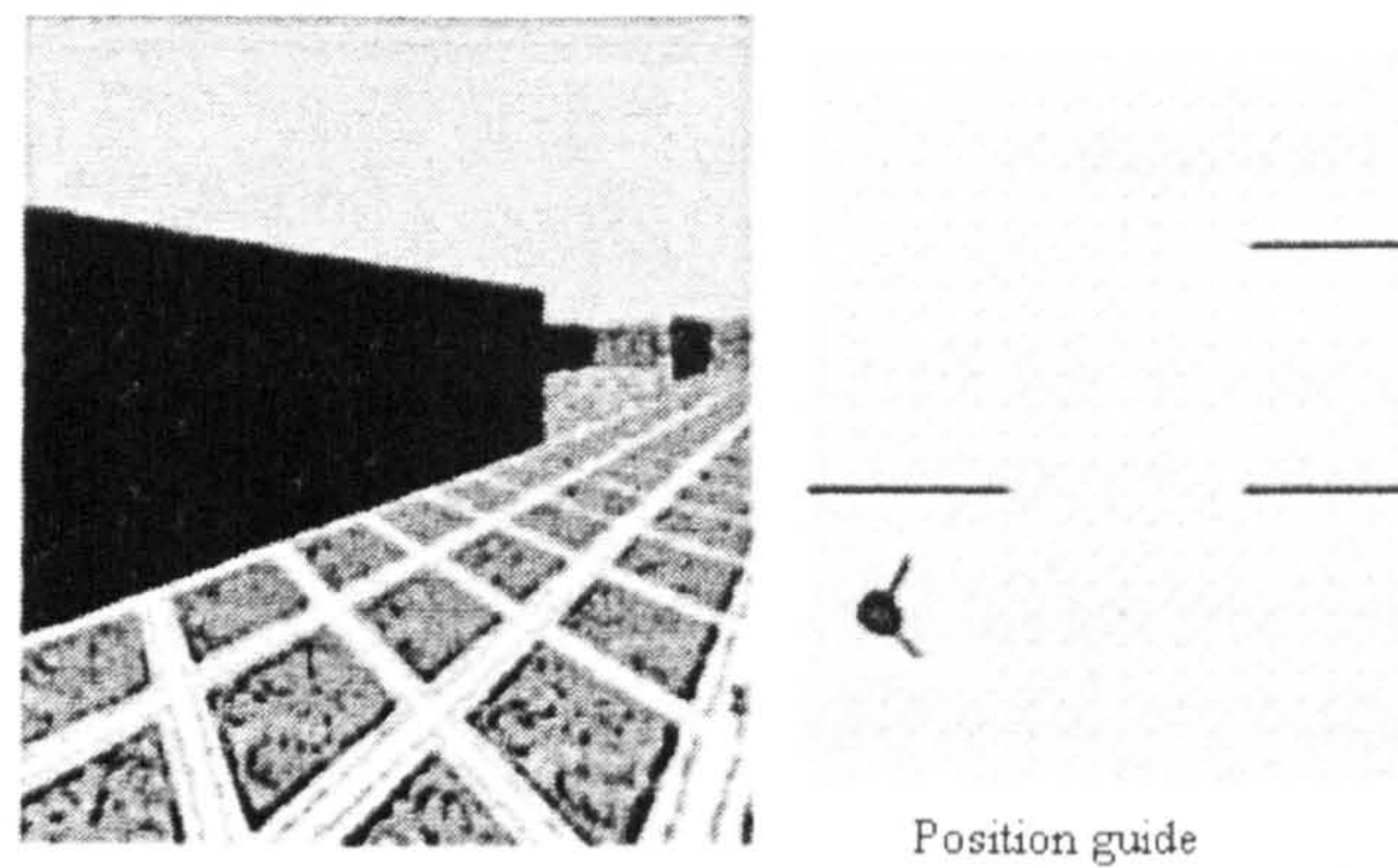


Figure 63: The field of view and the position guide in the 3D interface

3. Sense of 3D depth of space

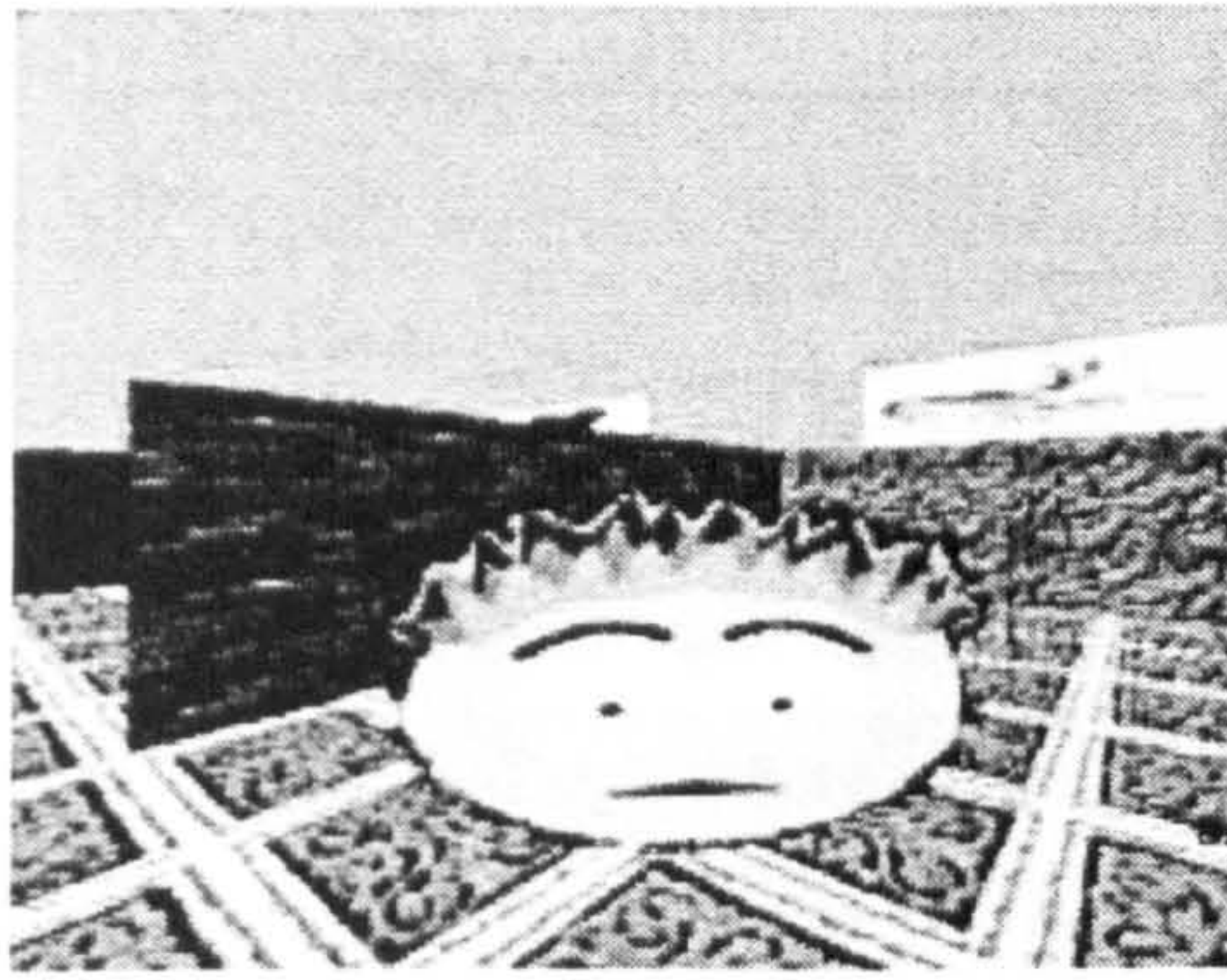
Space depth is a widely presented feature in both 3D games and in real life. Items far away from us will look smaller than their actual size, however most 2D interfaces cannot provide the same feelings. An example of the sense of depth can be found in figure 64. In figure 64, it can be seen that the far away wall looks much smaller in the viewing component, although the walls were designed to be the same length.

4. Exploration of space

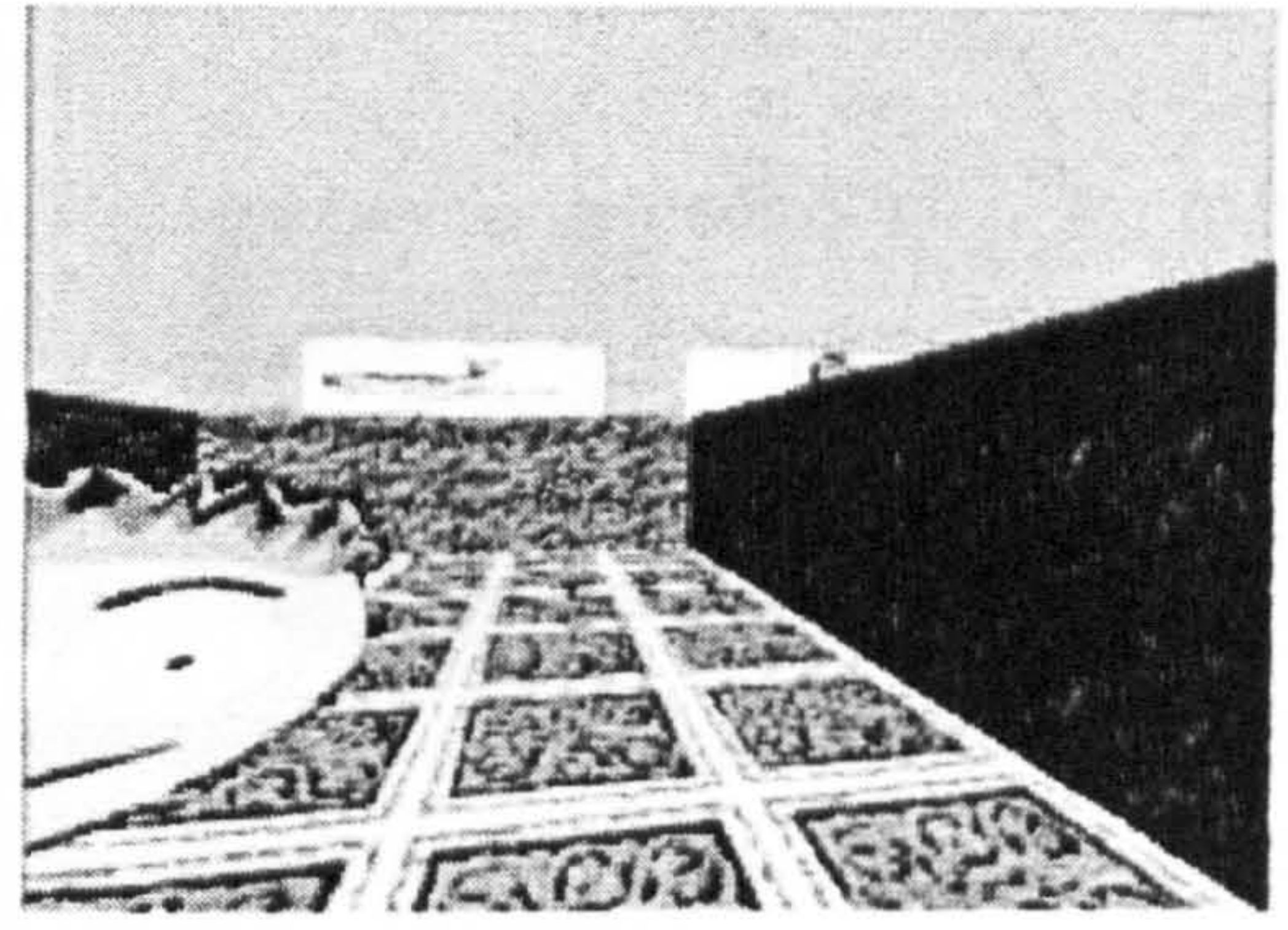
For the 2D interface, the whole interface is presented to users. Users know at the beginning who is in the system, whom are they talking to and what functions the interface provides. For the 3D interface, users need to explore the space to meet others or to access the assistant functions provided by the system. A position panel can only provide some aspects of the overall location and limited user information.

5. Eye contact

In real life, it is impolite to turn our backs on people talking to us. For the 2D text based interface, users can not move their positions and the positions of the images representing them are fixed. Thus no virtual eye contact can be established. The 3D interface provides the possibility of making virtual eye contact as users move around the 3D interface. Figure 64 demonstrates the viewing component of the 3D interface, which shows the direct glance and side glance.



(a) Direct glance

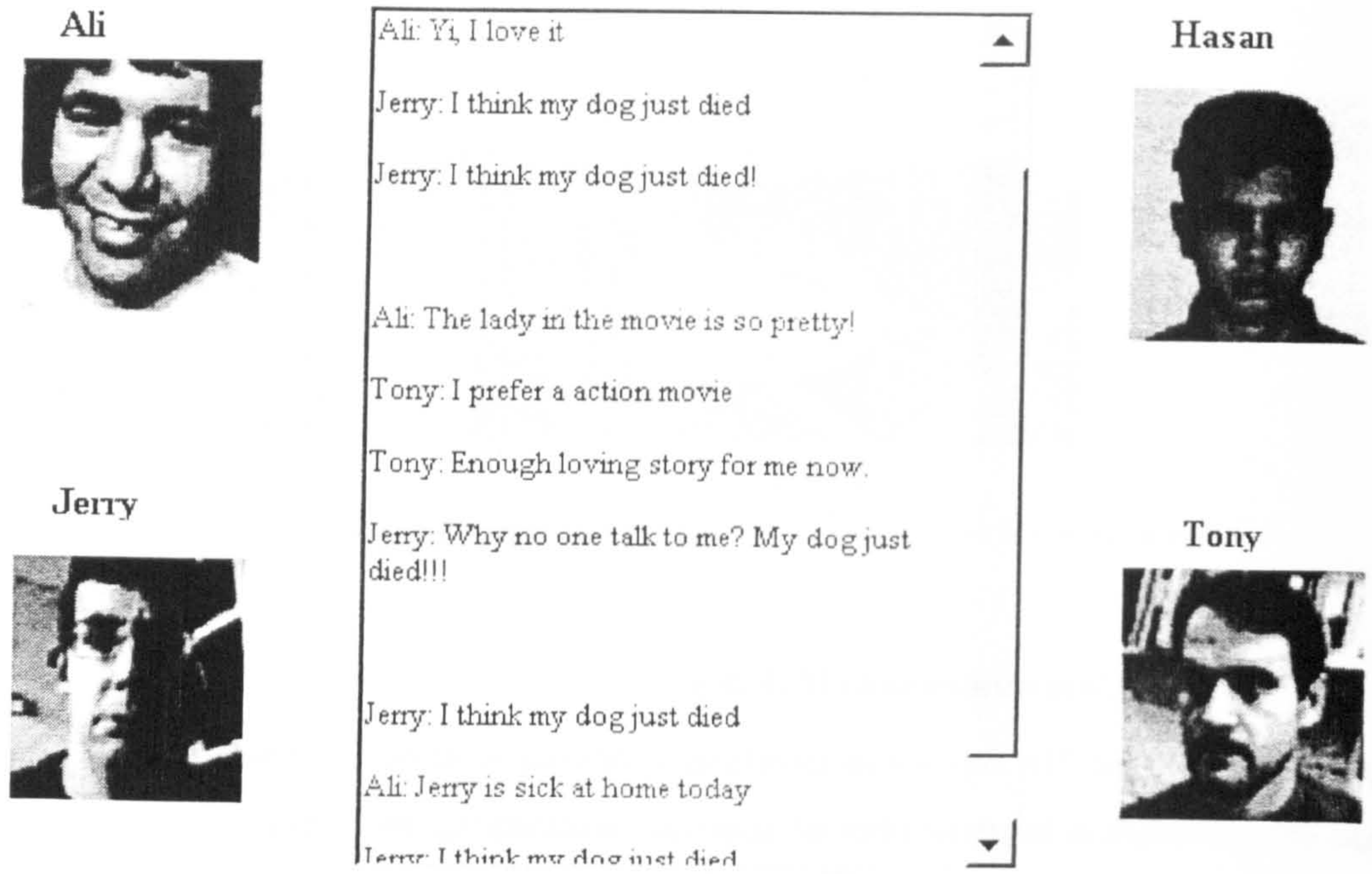


(b) Side glance

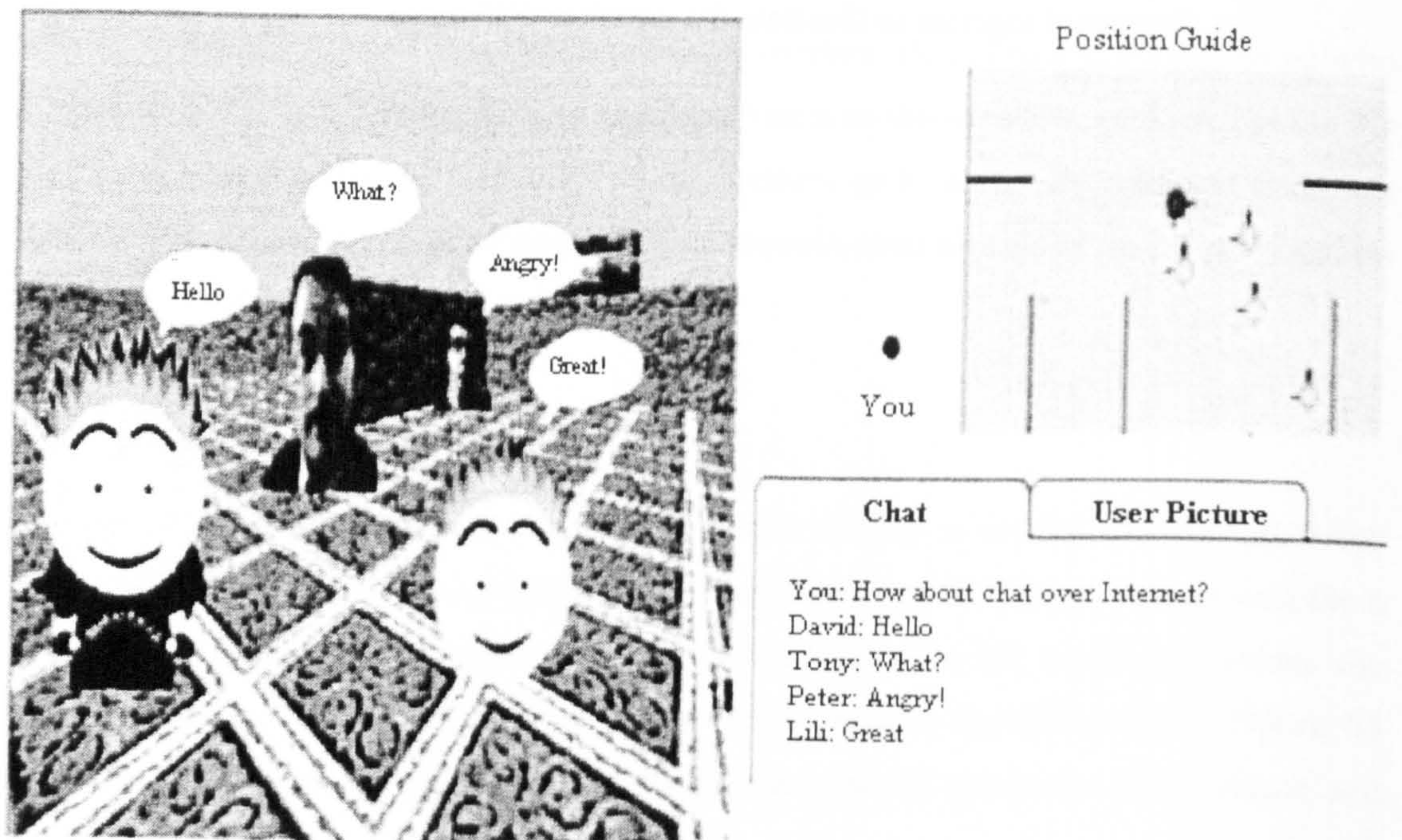
Figure 64: Different eye contact angle

6. Communication efficiency

In the 2D expressive interface, user's input sentences can be viewed by everyone. However when a large number of users are exchanging messages quickly, it is difficult to follow the messages of a particular user. For the 2D chat interface, communication will not be efficient when a large number of people gather in the same room. With the speech bubbles, users in a 3D expressive interface can identify other users' chat messages relatively easy. Users can concentrate on one user's speech by moving and changing the eye angles. Figure 51 and 52 in chapter 7 presented a busy 2D and a 3D chat environment. In figure 65, these two figures are presented together to illustrate the interface difference.



(a) A busy 2D chat interface



(b) A busy 3D chat interface

Figure 65: Busy 2D and 3D chat interfaces

7. Social attraction feeling

In daily life, when we find that some people gather together to discuss something, we may assume that interesting events or an urgent situation occurred. The social attraction feeling still applies to Internet communication environment. For the 2D interface, users can judge the number of participants in one discussion group by scrolling through the text. For the 3D interface, users can find this information visually by glancing for clusters of gathered users from the position panel and the viewing component. Figure 65 (b) demonstrated this feelings in a 3D interface.

8. Movement plus talk

In the 2D expressive interface, user's position is fixed. For the 3D interface, movement is a fundamental element. Users may move around to explore the world to find other users to chat with. It is also quite possible that some users may chat while moving round virtually.

The expressive 2D interface and 3D interface present different styles. Compared to the 3D interface, the 2D interface is a less sociable environment. As a less sociable environment may lead to less social presence feeling, the 2D interface may present less social presence than the 3D interface.

9.5 Method of the experiment

The first aim of this experiment was to examine the preference of participants for different interface styles when performing different types of activities. The second was to assess the effect that an individual's sociability undertakes on the satisfaction rating of each style of interface.

According to the purpose of the communication, there were two types of activities to be carried out in both interfaces, i.e., business-oriented activities and social-oriented activities. However, instead of actually carry out these activities in the 2D and 3D test platform, the activities were listed in a questionnaire and the experiment participants were asked to choose the preferred interface between the 2D and 3D interface for individual activities. The questionnaire contained questions such as "for activity - discuss stock market news, which interface do you prefer?". Experiment participants could choose the 2D interface, the 3D interface or neither of them. The choices of each participant were recorded for further analysis.

To assess the preferences of different interfaces for different activities, twelve activities that can be carried out in both two interfaces were created. Six of the questions belong to the business-oriented group and other six are classified into the social-oriented group. The activities are shown in table 38.

Activities	Style
Conduct a technical meeting over Internet. (business 1)	business-oriented
Seek technical advice about your computer. (business 2)	business-oriented
Monitor your employee's progress. (business 3)	business-oriented
Online study. (business 4)	business-oriented
Do your maths homework (e.g. $3x + 5y = 70$) (business 5)	business-oriented
Discuss stock market news (business 6)	business-oriented
Chat about the latest celebrity gossip. (social 1)	social-oriented
Seek new friends. (social 2)	social-oriented
Watch an animation. (social 3)	social-oriented
playing a multi-player game, e.g. football. (social 4)	social-oriented
A private chat with your good friends (social 5)	social-oriented
Display an exhibition of your paintings (social 6)	social-oriented

Table 38: The style of activities

To assess the sociability of each experiment participant, a questionnaire was adopted from Bellamy and Hanewicz's work [1999], which contained seven items with five scale points ranging from Agree to Disagree in order to measure the sociability of the participant. The questions are listed in table 39.

Questions 1	In my free time I like to interact with other people.
Questions 2	I prefer classes where the students get to work in groups.
Questions 3	I enjoy going to parties.
Questions 4	I enjoy being by myself most of the time.
Questions 5	I enjoy belonging to organisations (e.g., fraternity/sorority, church group, political group, etc.).
Questions 6	I enjoy meeting new people.
Questions 7	I am comfortable in new social situations.

Table 39: The sociability assessment questions

9.6 Experiment operation

A total of 50 students and staff from Bournemouth University participated in the experiment. First the gender of each participant was recorded and the questionnaire assessing the sociability of each participant was shown. After finishing this section, participants were redirected to view the two expressive interfaces (2D and 3D). After the viewing, participants were given twelve activities (in table 38) that can be performed in both environments. The participants were instructed to select the style of interface that was best suited for specific activities.

There were three main measurements that were recorded in this experiment. The first was the gender of each participant and the answer "male" or "female" were mapped to value 0 or 1. The second was the choice of interfaces for each activity. The answer "2D", "3D" or "neither" was mapped to values 1, 2 and 3. The participant's sociability was calculated by following Bellamy and Hanewicz's original methods [1998]. In the sociability questionnaire, answers' values vary from 1 to 5 corresponding disagree to agree. The higher the score that a participant achieved the more sociable the participant is. In Bellamy and Hanewicz's methods, participants who achieved a score higher than 25 were classified as high sociable and those who achieved a score lower than 25 were classified as low sociable.

The preference of interfaces was first examined by testing the hypothesis that the 2D interface would be more preferred for business-oriented activities and the 3D interface would be more preferred for social-oriented activities. Then the preference of interface was examined by testing the hypothesis that gender and sociability of the participants would influence the preference.

As gender and participant' sociability may influence the preference of interfaces, participants were divided into groups first according to their gender, then according to their sociability. Each group was compared against each other using a number of procedures, i.e., the independent-samples t-test and correlation test, to test the statistical significance of the observed preference differences.

9.7 Experiment results

9.7.1 Percentage analysis

According to the activity style, the collected results were classified into two categories. The first step of the analysis was to examine the percentage of interface preference to business-oriented and social-oriented activity styles. The results of the business-oriented activities are shown in figure 66, and figure 67 presents the results of the social-oriented activities.

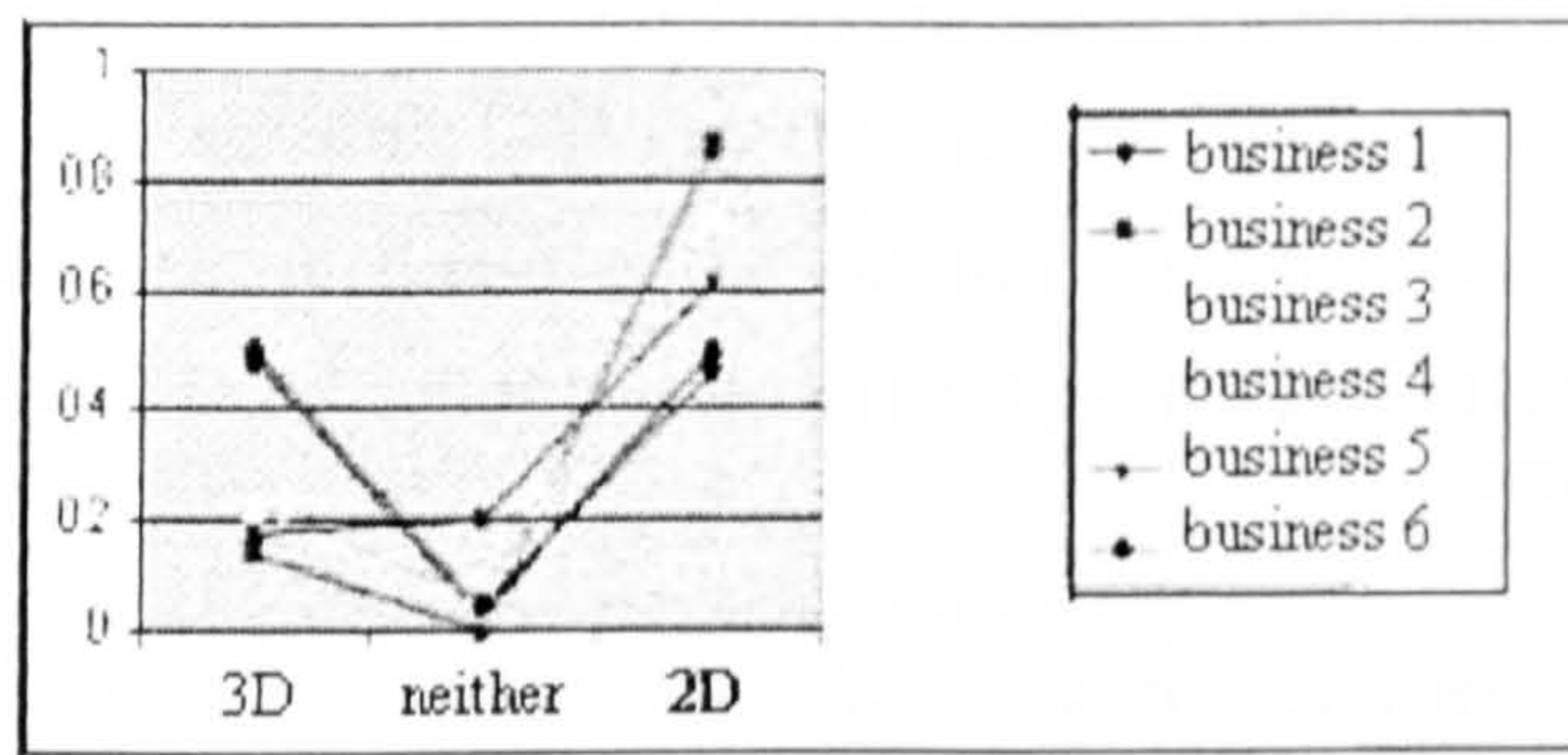


Figure 66: The percentage of interface preference of the business-oriented activities

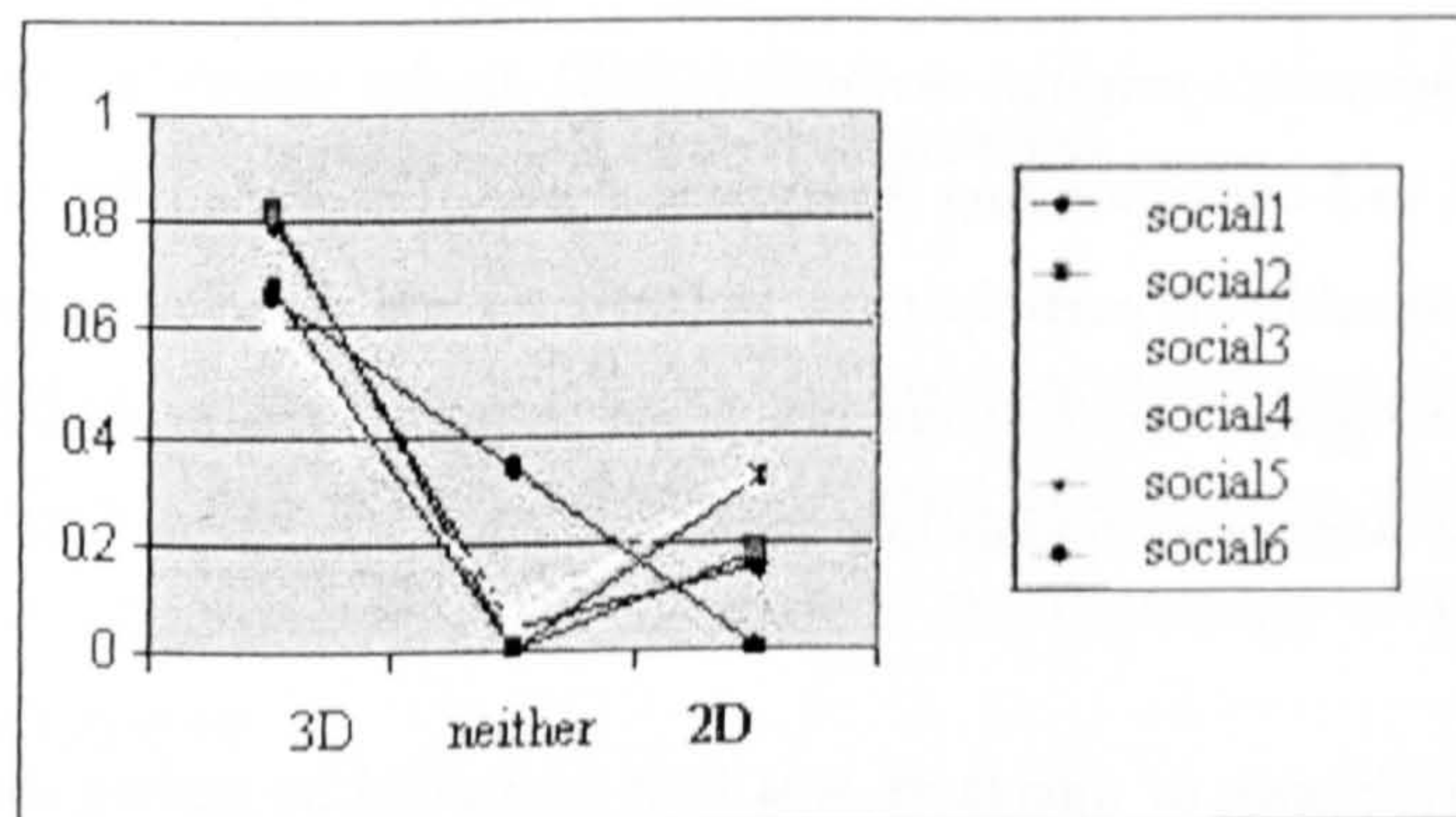


Figure 67: The percentage of interface preference of the social-oriented activities

Figures 66 and 67 show that more participants chose the 2D interface for the business-oriented activities and the 3D interface was chosen for most social-oriented activities.

However conclusions can not be drawn before testing the possible statistical significance. Table 40 lists the summary of the dependent variables (2 × 3) applied in statistical tests.

Activity style	Interface choice
business-oriented	2D interface
social-oriented	3D interface
	neither

Table 40: Dependent variables

9.7.2 Activity style Vs interface preference

The above percentage analysis demonstrated that business-oriented activities were more preferred in the 2D interface and social-oriented activities were more preferred in the 3D interface. To assess the influence of activity style to the interface preference, correlation tests were carried out to compare the relationships between the preference of interface and the activity style. The correlation test result is 0.597, which is significant (p=0.01).

The result demonstrated that there was a significant relationship between activity style and preference of interfaces. When users need to carry out a business-oriented activity, e.g. find an emergency telephone number, most people will prefer a simple interface that presents a low level of social feelings. When users want to perform a leisure activity, e.g., playing an online game, the preferred interface is a relative complex and vivid multi-channel environment.

The practical hint for interface design is: if the purpose of the online communication is to provide technical help or technical discussions for the users, a simple, straightforward and uncluttered environment is the best choice. If the purpose of the online communication is to relax and to enjoy the online lifestyle, a vivid, video or audio assisted environment will be preferred by most online surfers.

9.7.3 Sociability Vs interface preference

It was expected that participant's sociability would influence the choice of interfaces. Correlation tests were carried out, however no significant link between the individual's sociability and interface preference were detected. This may indicate that participant's sociability is not significantly related to their interface preference. However, when sociability is treated as a dependent variable, will the interface preference differ between the low sociability and high groups? A t-test was carried out to compare the ratings of the two groups and the results showed that the distribution of preference ratings of high sociable participants was not

significantly different from the less sociable participants for the twelve individual activities. However, a marginally significant difference was found in the t-test results between the two groups of the participants for business-oriented activity 1 (“conduct a technical meeting over Internet”) and social-oriented activity 2 (“seek new friends”), in which $p = 0.08$ and $p=0.09$ correspondingly.

To further explore the influence of sociability, the mean values of the twelve interface choices were calculated and a t-test was carried out to assess whether the mean ratings of all the activities differ between the sociable and less sociable groups. The results showed that there was a marginally significant difference between ratings of the sociable and less sociable participants ($p=0.06$) for the mean value of the twelve choices.

The experiment results demonstrated that in the overall level, sociability has a marginally significant influence on preference of interface style. On average, low sociability persons preferred simple and clean interfaces and high sociability persons preferred complex and realistic interfaces. However, for specific activities, human sociability had limited influence on interface style preference.

The influence of an individual's sociability was much weaker comparing to the significant influence of activity style. This provides another design criterion for online communication. Designers should focus more on the activities that will be carried out within their systems. However if a series of online communication interfaces will be presented to a specific user, the user's sociability should be considered and designers should adopt sociability into the design consideration.

9.7.4 Gender Vs the interface preference

Are there any differences of interface preference between genders? Will a female prefer a vivid online environment more than a male? To answer this question, the t-test and correlation test were carried out. Correlation test showed no significant correlation and t-test result showed one marginally significant difference ($p=0.075$) and one significant difference ($p=0.04$) between the ratings of males and females, which was found at the business-oriented activity 2 (“seek technical advice about your computer”) and at the social-oriented activity 3 (“Watch an animation”) correspondingly. The mean value of all the activities was calculated and a t-test was carried out. The test result did not show any significant preference difference. This indicated that gender had a very limited amount of influence on the preference of interfaces.

9.7.5 The analysis of the hypotheses

In the hypotheses, it was argued that activity style may strongly influence the preference of interface; sociability may strongly influence the preference at an overall level and that interface preference would not be strongly influenced by differences in gender.

From the experiment result analysis, it can be seen that activity style had a strong influence on the preference of the interface. For social-oriented activities, participants preferred the 3D expressive interface presenting more social presence feelings. For business-oriented activities, the preference choice was opposite. No significant influences were found when comparing the sociability style with activity styles. By calculating the mean choice values for all the activities, the analysis demonstrated that an individual's sociability strongly influenced their interface preference at a general level. The first and the third hypotheses were supported by the experiment results, while the hypothesis 2 was not supported.

9.8 Chapter conclusions

This chapter presented an experiment that assessed the factors related to preference difference between the 2D and 3D expressive interfaces. Three main factors - sociability, activity style and interface style were examined. Different interface styles present the different amounts of social presence feelings to users. The more social presence feelings presented the more realistic and more sensible the environment is. An individual's sociability may influence their life style and their preference of environment. The activity style is categorised by the purpose of the communication and different purposes may lead to different environment preferences.

The feature comparisons between the 2D and 3D expressive interfaces illustrate that the 2D expressive interface is a more straightforward interface and the 3D interface provides more aspects of virtual reality, which are more complex. The experiment results showed that significant differences existed between the preferences of interfaces for different activities. A business-oriented activity may be more suited to the 2D expressive interface, which is a straightforward environment and a social-oriented activity may be more suited to the 3D expressive interface, which leads to a vivid virtual reality interface. The individual's sociability may also influence their preference of interfaces. However, a significant difference in preference was only revealed at the overall level, and not for the majority activities. This indicated that an individual's sociability did have impacts on their preference of the expressive interfaces. However, when dealing with specific activities, other factors such as activity style, may influence preference much stronger than the individual's sociability. Gender does influence

social preference in some specific activities, but no significant preference differences can be found for the majority of activities at the overall level.

This chapter and the previous chapter described experiments assessing the preference of interfaces with automatic expressive image display. The experiments' results demonstrated that not only users prefer an interface with expressive image display but also the preference will differ when users need to carry out different activities. In next chapter, the focus will change from assessing the preference of interfaces with automatic expressive image display to assessing the influences of automatic expressive image display. Experiments assessing the influences of the expressive image display on human-computer-interaction and the factors may affect the influences will be presented.

Chapter 10

The influence of expressive images display on HCI and the factors affecting the influences

In chapter 8, the experiment results showed that users prefer interfaces with expressive image display. The previous chapter demonstrated that people intend to carry out different activities in the 2D and 3D expressive interfaces. However what kind of influences does the automatic expressive image display have? In this chapter, an experiment assessing the effects of displaying expressive images to users is presented. Then the factors that may affect the influence of expressive image display are discussed and an experiment assessing two of these factors is described.

10.1 The experiment assessing the influence of expressive image display

This chapter is divided into two sections. In this section, the experiment assessing the influence of the expressive image display is discussed. Firstly, the background considerations of the experiment, i.e., social norms, emotion, and AI agents, are discussed. Secondly, the problem that the experiment was attempting to solve, the aims of the experiment and the experiment method are presented. Thirdly, the implementation of the experiment, including the experiment interface design, task description and the expected experiments results are shown.

10.1.1 Background

10.1.1.1 Social norms

A norm is a generalised disposition to the world shared by members of a community [Liu 2000 and Ronald et al. 2000]. A social norm can be defined as a generally accepted way of thinking, feeling, or behaving that people in a group agree on and endorse as right and proper [Brehm et al. 1999]. Social norms direct our behaviour and guide our feelings. The influences of social norms for daily behaviour include conformity and compliance [Brehm et al. 1999].

Emotion and social norms are related. Imagine you are in a happy mood and start to sing. You are not a good singer but it does not matter now since you think you are alone. However, soon you realise that you are not alone and some young girls near a corner are looking at you and

begin to laugh. You feel the tendency to hide, which is a sign of being shamed. The above example suggests the emotion such as shame plays an important role in sustaining social norms. Elster [1996 and 1999] defines social norms as injunctions to behaviour with the following features:

- First, social norms define the expectations about what behaviour, thoughts, or feelings are appropriate within a given group within a given context. Social norms are "Do something", "Do not do something", or "if others do something, then do this" if the social norm is conditional. In this case social norms are not outcome-oriented or future-oriented.
- Second, social norm theory assumes that behaviour is influenced by how other members of our social groups behave. As the name suggested, social norms must be shared by other people and people are important for enforcing them through sanctions.
- Third, social norms are not only sustained by the sanctions of others, but also by emotions. People are supposed to have appropriate emotions under different circumstances. For example, people should be grief and sorrow when attending a funeral; as a consequence happiness is inappropriate.

Elster made the assumption that social norms play an important role in the generation of emotions such as contempt and shame. In addition, he noted that emotions and their expression may be regulated by social norms [Elster 1996].

10.1.1.2 Emotions and social norms

Social norms enter the process of emotion generation during appraisal. Many emotions were contingent upon adherence or violation of social norms [Elster 1996]. Scherer [1988] provided a table of the complete appraisal patterns for some major emotions including shame, guilt, anger, contempt, and pride.

Social norms are crucial for the instigation of emotion regulation. Signals of external or internal outcomes of unrestrained emotional behaviour instigate regulatory processes [Staller and Petta 2000]. Early theories of emotion (e.g. Darwin's emotion theory) denied the relationship between social norm and emotion by stating that emotion was purely biological phenomena. Current researchers integrate emotion with social norms and argue that people do not present emotion simply because the animal within responds instinctively to certain predicaments [Heise and Calhan 1995]. Rather, human society forms the regulations intelligently to present their feelings and their displays of emotions.

10.1.1.3 Social norms and computer communication

Researchers started to analyse social feelings and regulations in human-computer communication and online communications when computers and the Internet emerged into our daily lives. From the 1960s, various researchers, e.g. Burns et al. [1961] and Hilitz and Turoff [1978], have proposed different theories and examined different aspects of emotion and social norms. At that time, the focus was on whether communicating with computers was a social interacting procedure.

By the 1990s, researchers, e.g. Steinfield [1985] and Brousseau [1990], had concentrated on explaining the social feelings that were presented in human-computer interaction and computer mediated communications (CMC). Different theories were presented to explain the online social norms. The most well-known theories include:

1. **Deficiency.** Turkle [1984], Winograd and Flores [1987], Dennett [1988] and other researchers argued that persons who respond socially to computers are either young, ignorant or socio-emotionally imitated. People apply social rules because no other rules can be applied [Nass and Sundar 1994].
2. **Parasocial Interaction.** Horton and Wohl [1956], Rubin and Perse [1987] and other researchers observed the parasocial interaction. In this theory, people interact socially with machines because they behave as if they are interacting with the person behind the machine. One example of this theory is that people treat software humanly because they are actually interacting with the programmers [Nass and Sundar 1994].
3. **Social Interaction.** This theory stated that the social feelings in human-machine communication are natural. In contrast to the parasocial explanation, a series of papers from Nass [Nass et al. 1994] argued that human-computer interaction was unmediated and directly social. Individuals respond to computers as a source in much the same way that individuals respond to other human beings as a source [Nass and Sundar 1994].

10.1.1.4 Emotional agent

In the computer world, agents may refer to a process, a program, a hardware controller or even a network component. In this research work, Franklin's definition of an agent is followed: "An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future" [Franklin and Graesser 1996].

An emotional agent is a computer system (e.g. a robot), probably including actuators and is structured in deliberative (or cognitive), reactive, and emotional modules, therefore able to interact with the external world, including humans and other agents. Agents can cooperate or compete with each other to reach their goals [Camurri et al. 1997]. The relationships between different agents are similar to the human world: they can be friends, colleagues, enemies or just strangers etc. An overview of an emotional agent structure adopted from the paper of Camurri [Camurri et al. 1997] is presented in figure 68.

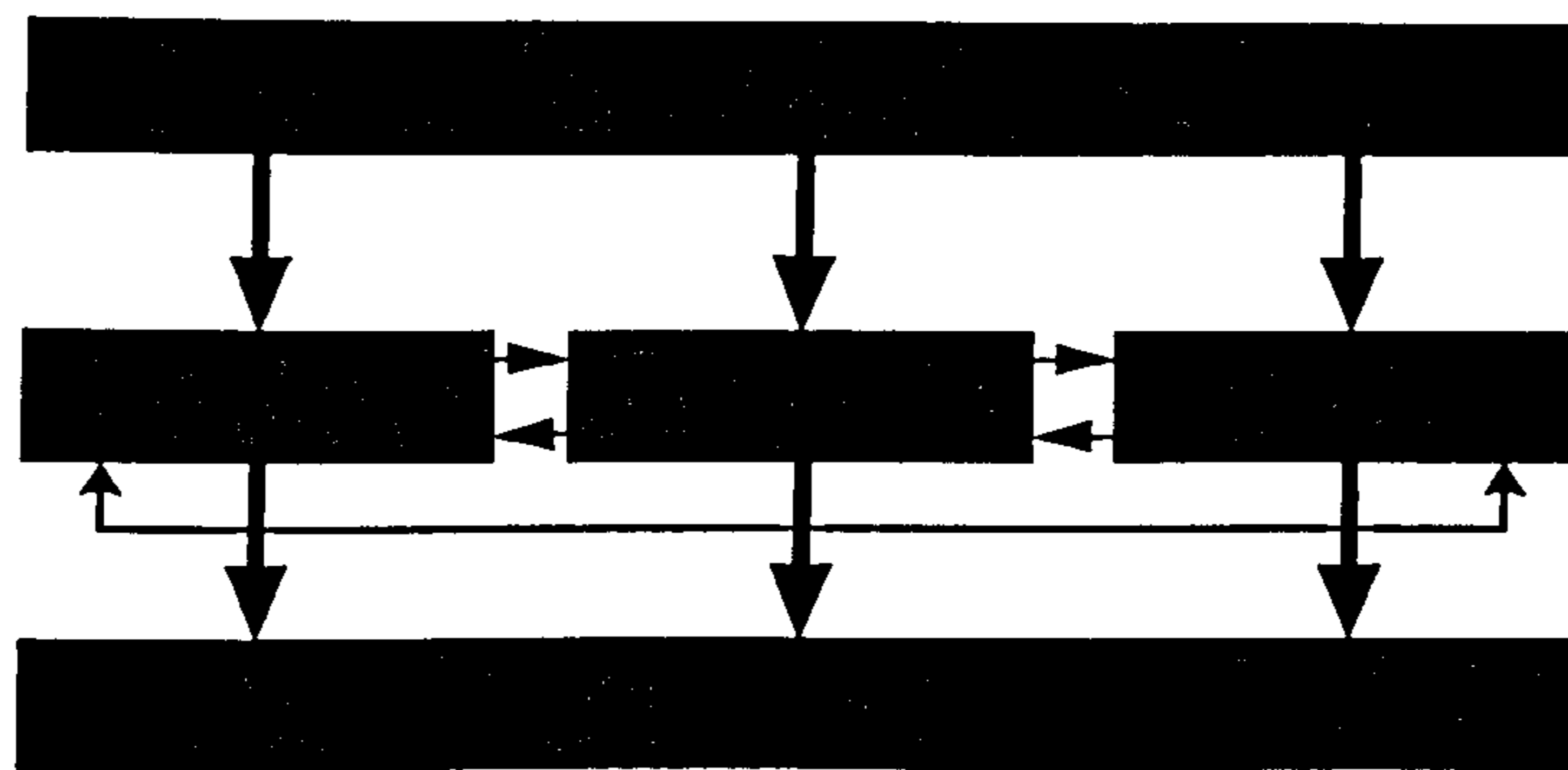


Figure 68: Structure overview of an emotional agent

The *input* component of an emotional agent varies from movement sensor to data transmitted by other agents via network. The *output* component sends out various signals, e.g. audio, video, text message or behaviour. The *reactive* component simulates the reflection behaviour of human beings. The *rational* component holds agent's rational state and external world information such as social norms, and geography information etc. The *emotional* component contains the agent's emotion state and sends or receives stimulus to or from other components. The *emotional* component may strongly influence the behaviour of other components and the output data. For human-agent interaction, especially for the human and emotional agent communication, the emotion expressions are extremely important. The emotional agent may show its expressions according to the internal emotion models, and then the observers, e.g. human beings, perceive the emotions.

10.1.2 Planning the experiment

From the above section, it can be seen that social norms undertake an important role in our daily life and Internet communications. When emotion expression is involved in the interaction, social norms may be strongly influenced. Following example shows such a scenario: You handed in your maths assignment. The teacher told you that you did a good job. Although you already self-judged that your assignment was excellent, the teacher's comments were perceived as more accurate than your judgement. The social norm behind is "other praise is more accurate

than self". However, if the teacher expressed a happy expression and gave a positive comment, you may feel greater intensity. On the other hand, if the teacher expressed a sad or disgusted expression and said "You did an excellent job", you might question the real meaning of the comment and even generate an opposite feeling.

To assess the influence of expressive images on Internet communication, a test environment needs to be set up in which the emotion expression displays are controlled both in categories and intensities. The presented emotion expressions should be a component naturally fitted into the scenario. Numerous Internet communication interfaces were investigated (e.g. chat room, online interactive game etc.) and the online human computer agent interaction interface, i.e., an online quiz interface, was chosen to be the test platform. The reason was that in an online quiz interface, the interaction contents (i.e. the question and answer session) could be fixed to every experiment participant and the emotion expression display could be naturally linked to the scenario (e.g. the agent may smile when the experiment participant chose a correct answer).

There were two main aims of the experiment. The first aim was to examine the social norms within human-agent communication and the second was to assess the effects that emotion expression display had on the experiment participants. It is impossible to examine all the social norms in our daily life or in Internet communication. In this experiment, social norm "*the independent judgement is fairer than self judgement*" was examined. It was expected that the social norm "*the independent judgement is fairer than self judgement*" was valid in human-agent communication. It was also expected that an emotion expression display correctly reflecting the internal feelings can strongly improve the perceived performances of the experiment participants. In contrast, an expression display opposite to the meanings may strongly decrease the perceived performances of the participants.

10.1.3 Methods of the experiment

To create the online human computer agent quiz platform, programming languages such as flash, JSP and HTML were used. The platform included an online quiz session and an online questionnaire session. Experiment participants can use their web browsers to open the quiz web page and carry out the online experiment.

10.1.3.1 The online quiz

The online quiz presented either one (agent Jerry) or two (agent Jerry and independent judge David) agents to the participants. Agent Jerry presented a small introduction and then asked

nine questions. Each question contained four choices. Agent Jerry indicated that the nine questions were randomly chosen from a database that contains 3000 questions. Actually, all participants received the same nine questions (see table 41). All the questions were designed to have an objective answer, instead of a participative one. Answers to the questions are not known or all possible options are equally likely. In the two-agent situations, participants were told that Judge David would check users' answers (independent valuation), while in the one-agent situation, agent Jerry would carry out this task (self-valuation). No matter what answers each participant provided, all participants were told that they answered the same five out of nine questions correctly. This ensured that all participants performed equally well.

When only one agent (agent Jerry) was presented in the experiment, the agent would present neutral emotions throughout the questionnaire session. When two agents (Jerry and judge David) were presented in the experiment, three interfaces were created:

- **Neutral expression interface.** The judge agent David showed neutral expressions for both correct answers and wrong answers of the participants.
- **Compliment expression interface.** The judge agent David showed happy expression for participants' correct answers and sad expressions for wrong answers.
- **Opposite expression interface.** The judge agent David showed happy expression for participants' wrong answers and sad for correct answers.

In summary, there are four interfaces (one agent, two agents with neutral expressions, two agents with complimentary expressions and two agents with opposite expressions). The typical screens of the four interfaces are illustrated from figure 69 to figure 72. In table 41, the questions and the answer choices are presented. The four categories that the online quiz interfaces classified into are presented in table 42.

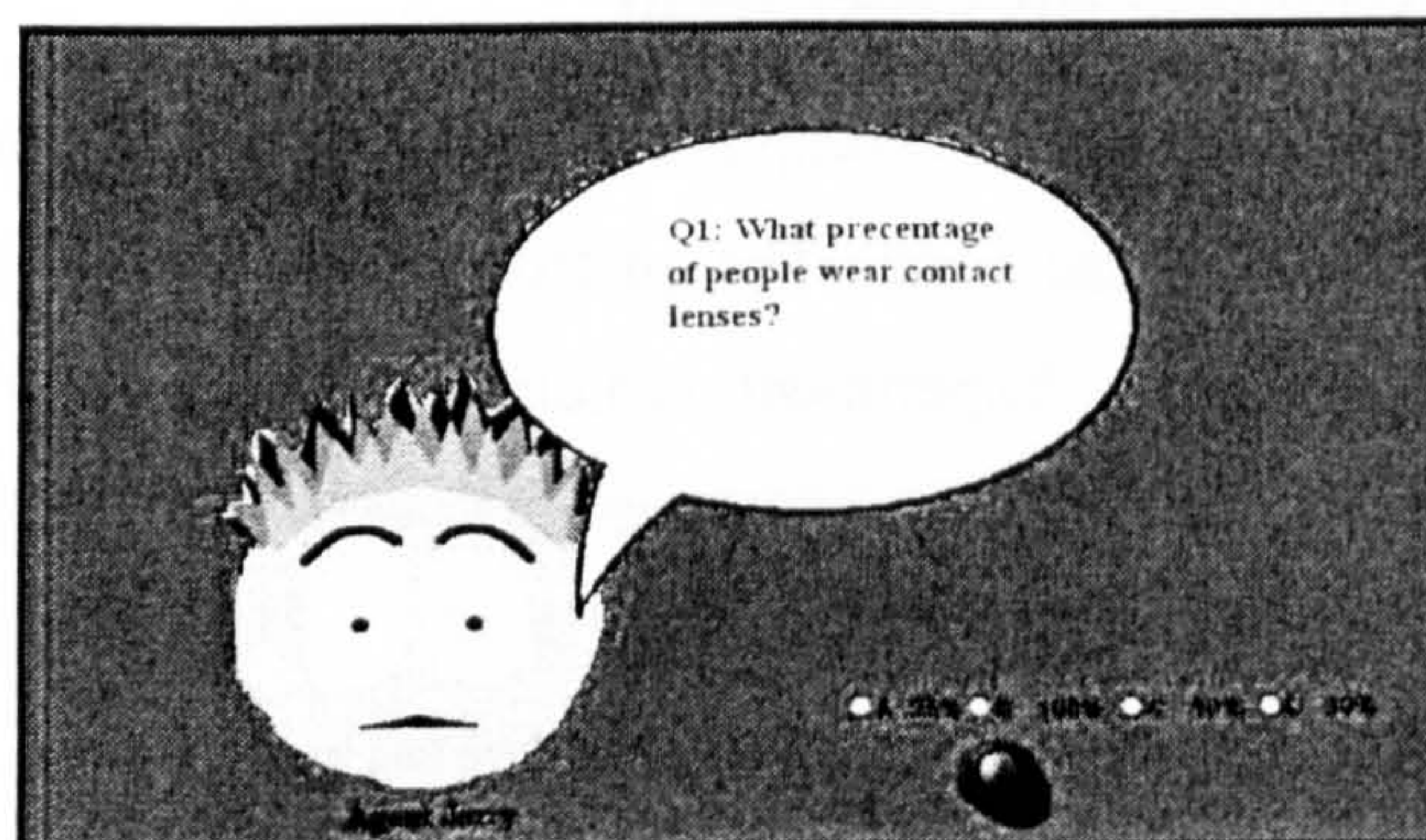


Figure 69: A screen shot of interface 1 (one agent)

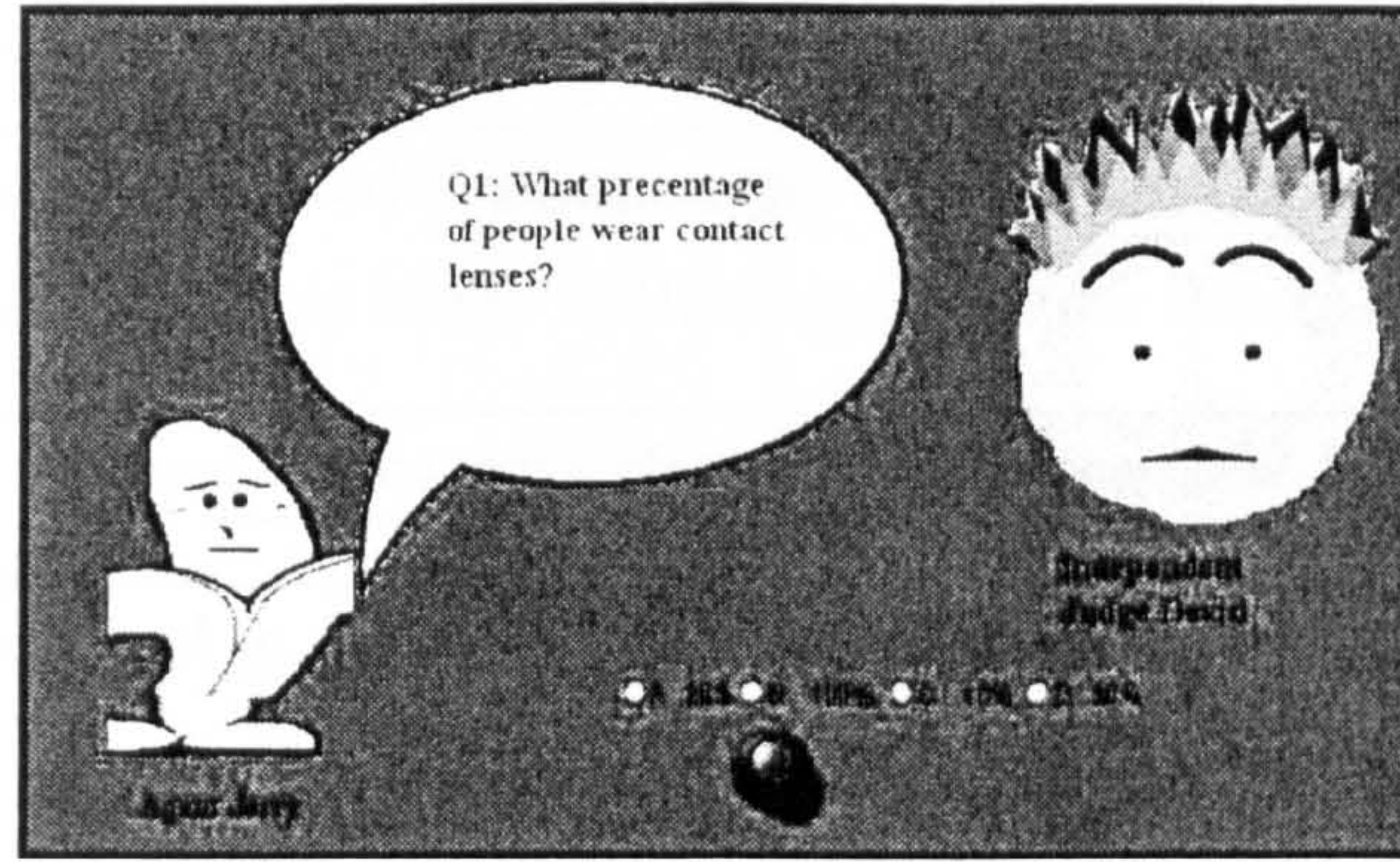


Figure 70: A screen shot of interface 2 (two agents with neutral expressions)

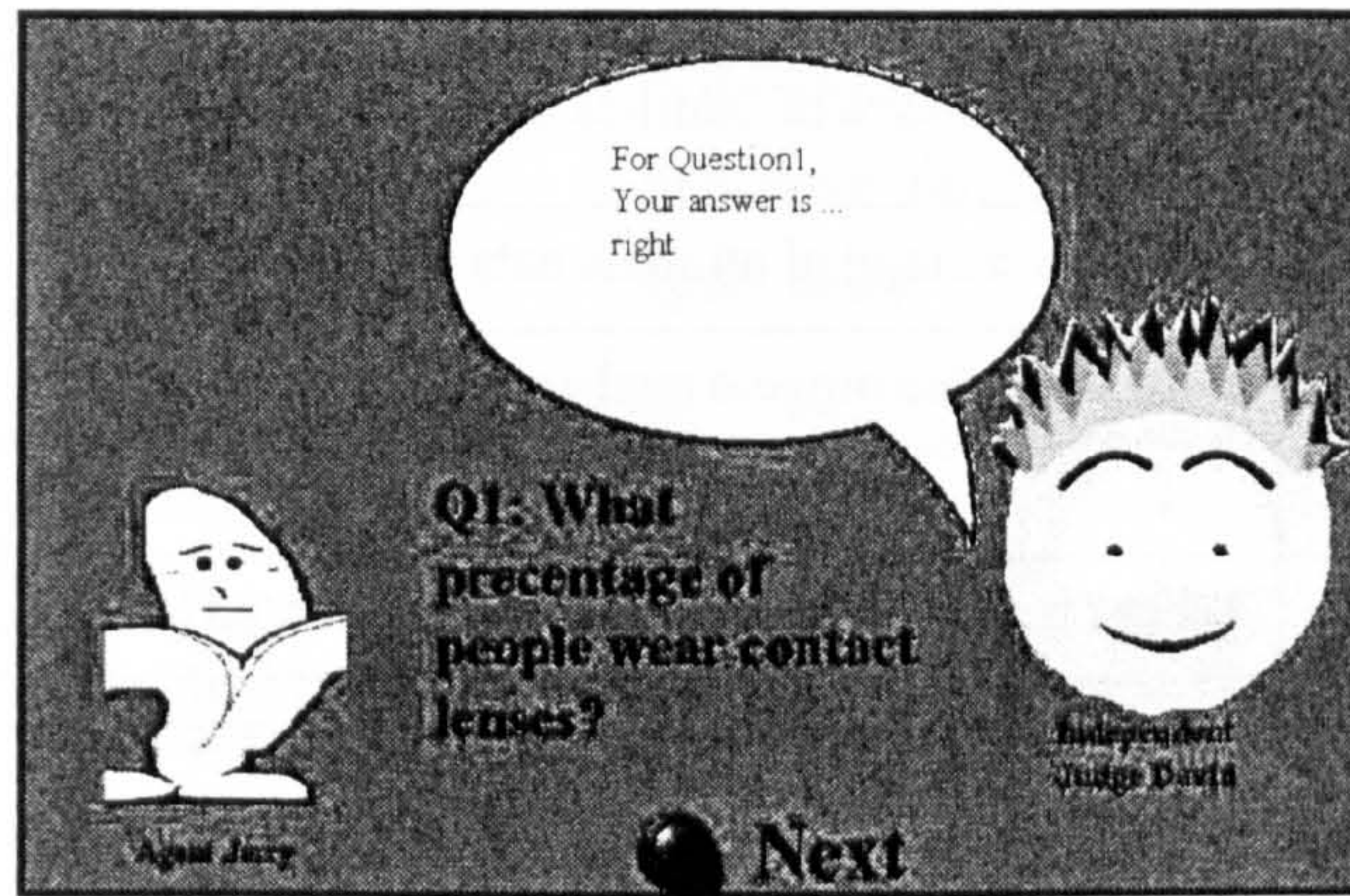


Figure 71: A screen shot of interface 3 (two agents with complimentary expressions)

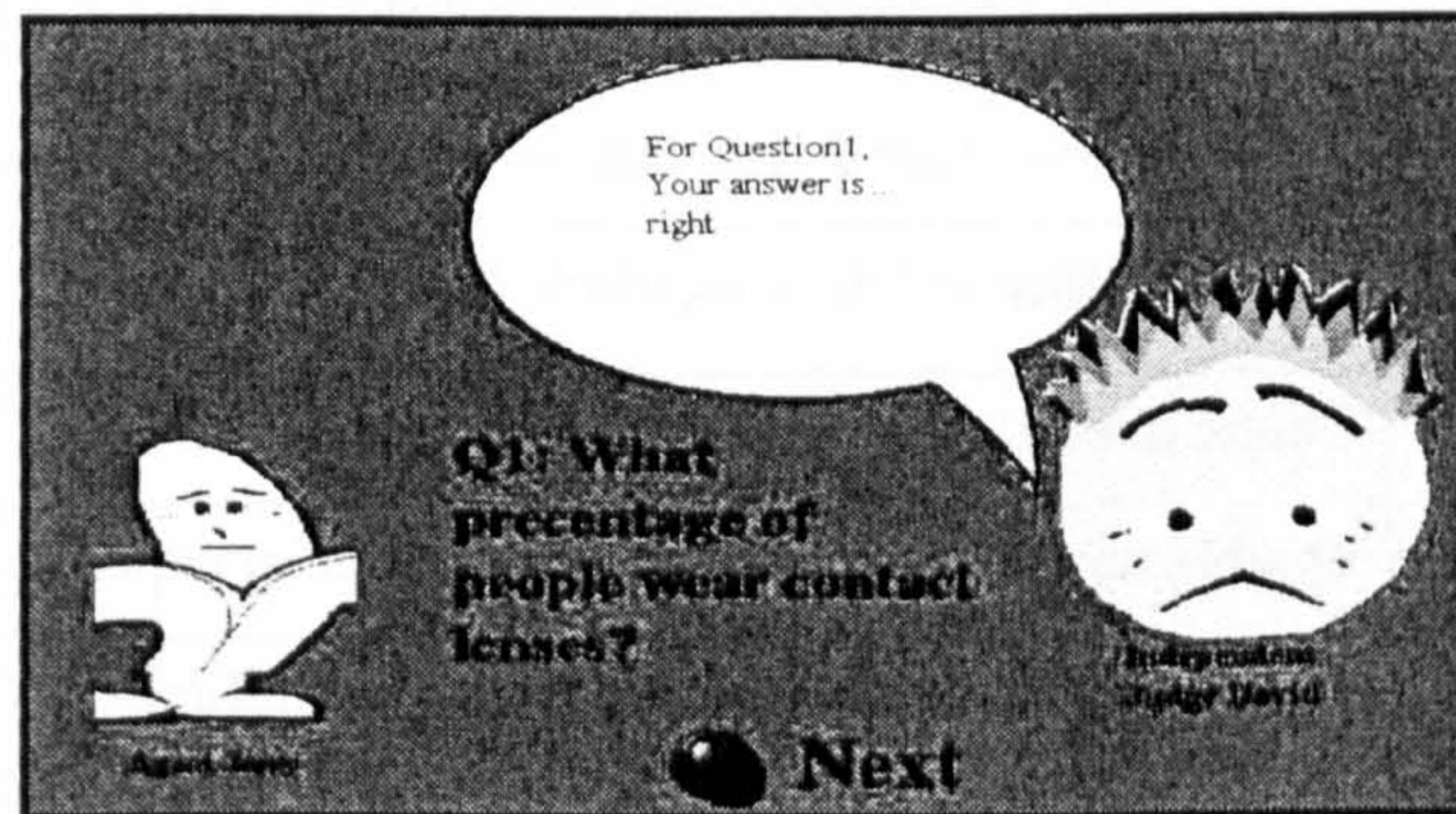


Figure 72: A screen shot of interface 4 (two agents with opposite expressions)

Question number	Questions	Answer choices
1	What percentage of people wears contact lenses?	25%, 100%,10%, 30%
2	What percentage of people consumes at least one type of organic food?	23%, 40%, 10%, 5%
3	What percentage of people tips less than 15% at a restaurant?	15%, 100%, 20%, 0%
4	What percentage of families owns a car?	33%, 47%, 50%, 65%
5	What percentage of animals eats meat?	13%, 37%, 20%, 75%
6	Who made the biggest apple pie in the world?	Mac, Burger king, A Chinese take away, George Bush
7	Where is the island "Jerry's Dream" located?	UK, USA, China, South Africa
8	Where can we find the most expensive computer?	NASA, BT, FBI headquarter, MIT computer lab
9	Which animal represents "good luck" in Chinese tradition?	Pig, Dog, Snake, Rat

Table 41: Questions and the choices for the online quiz

Interfaces	Description
1	Single neutral agent without facial expression
2	Double neutral agents, judge without facial expression
3	Double emotional agents, judge with complimentary facial expressions
4	Double emotional agents, judge with opposite facial expressions

Table 42: Four categories of the online quiz interfaces

In the online quiz session, participants' answers were not recorded as the interface only intended to simulate a quiz environment and the answers in this session will not influence the social norms assessment. To measure the social norm *"independent judgement is fairer than self judgement"* and the emotion expression's role in human-agent interaction, a questionnaire was devised.

10.1.3.2 Questionnaire

The questionnaire session included five questions related to the perceived performance of software agents and the online quiz itself. Table 43 presents the five questions.

Questions	Explanations
1. How fair the judge is?	For single agent interface, it refers to agent Jerry, for double agents interfaces, it refers to agent David.
2. How friendly is the judge to you?	Is the judge nice to you?
3. How friendly is the judge to agent Jerry?	Is the judge nice to agent Jerry in the double agent interfaces (this question is not applicable for single agent case)
4. How is your perceived test performance?	Do you think you answered the questions fairly well?
5. How easy the quiz is?	No explanations

Table 43: The questions in questionnaire session

10.1.4 Experiment environment

Fifty students and staff volunteered to participate in this experiment. All the participants were experienced with computer manipulation or Internet surfing. These participants were quite knowledgeable concerning computers so that the deficiency problem (people treat computer as humans because that they do not have enough experiences with computer) described by Nass and Sundar [1994] does not apply.

The participants performed the experiment through the Internet using their web browser (IE or Netscape). When the participants opened the first page of the experiment web site, some introduction and welcoming messages were given. Then, the participants were randomly directed to one of the four possible interfaces and interacted with the agent(s) to finish the quiz. When the quiz was finished, the participants were redirected to answer the questionnaire. The entire experiment took approximately 15 minutes.

Out of the fifty participants, twelve viewed the one agent interface, thirteen viewed the two agents with neutral expressions interface, eleven viewed the two agents with complimentary expressions interface and the remaining fourteen viewed the two agents with opposite expressions (see figure 70 -73 for interface details).

10.1.5 How the data was measured and analysed

All the data collected in this experiment were within the questionnaire session. The online quiz session aimed to create an atmosphere so that participants thought that the agent was judging according to their answers.

There were three main measurements that were recorded in this experiment. The first was the gender of each participant and the answer "male" or "female" were given value 0 or 1. The second were the answers of the questionnaire and the values vary from 1 to 10, which 1 means lowest, 10 means highest. The third was participant's sociability, which was calculated by following the same procedure described in the previous chapter.

The perceived feelings of interfaces were first examined by comparing the feedback from different interfaces. As gender and participant's sociability may influence individual's perceived feelings, participants were divided into groups first according to gender, then according to sociability.

10.1.6 Experiment results

Question 1: The fairness of the judge

The mean participants' ratings (out of 10) for the question "the fairness of the judge" for the four different interfaces and the standard deviations are shown in table 44. To view the ratings clearly, a graph was drawn in figure 73.

	Interface 1	Interface 2	Interface 3	Interface 4
Mean rating	6.09	7.72	7.57	5.5
Standard deviation	2.4	1.9	1.6	2.5

Table 44: Mean ratings and standard deviations for question 1

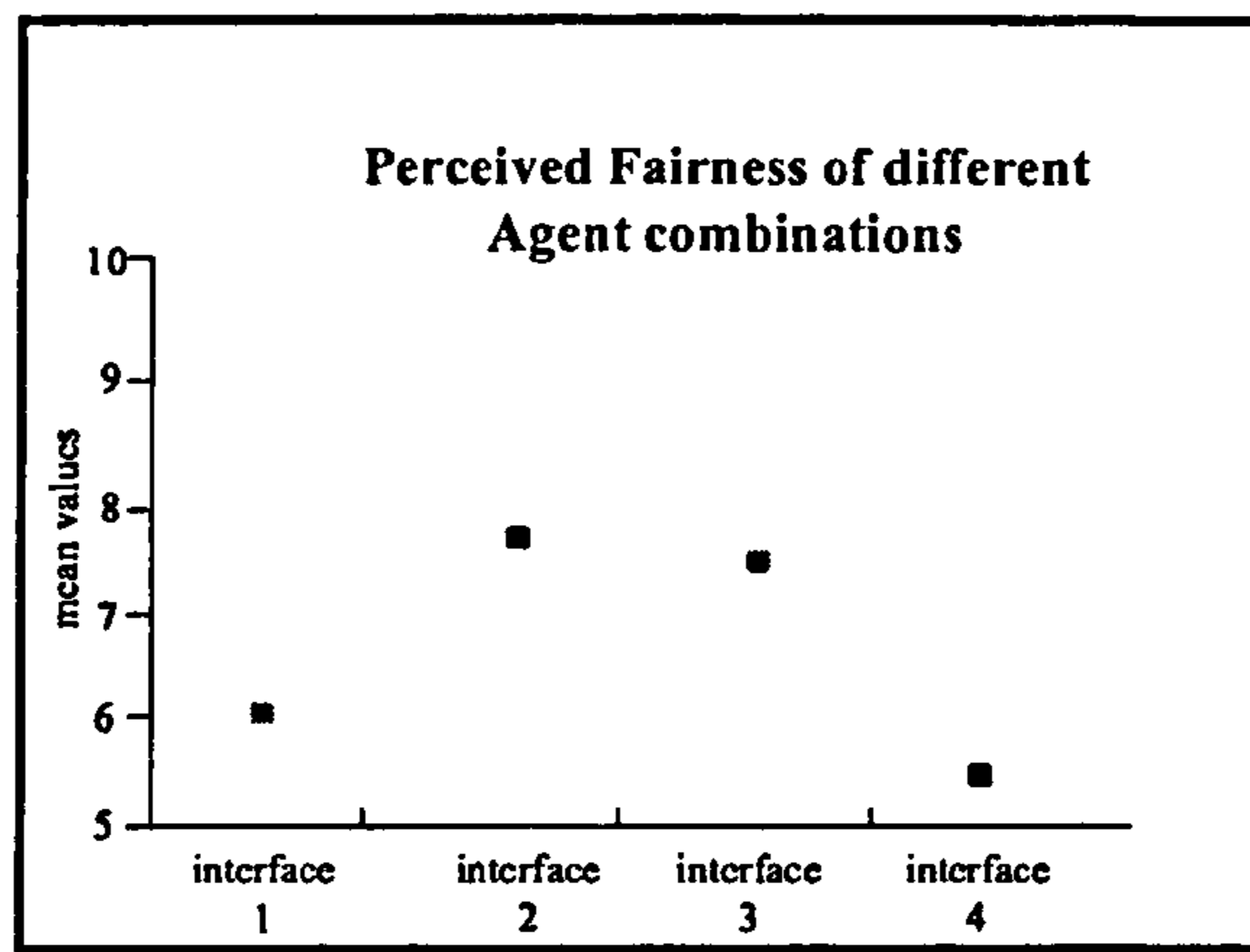


Figure 73: Participants mean ratings of question 1

It can be seen clearly that there were strong increases of mean ratings from interface 1 to interface 3 and from interface 1 to interface 2. Other strong increases of mean ratings can be found from interface 4 to interface 2 and from 4 to 3. However, conclusions can not be made before statistical analysis is carried out. To compare the mean rating in different interfaces, t-test was carried out and a marginally significant difference was found when comparing the ratings of participants who viewed interface 1 with the ratings of participants who viewed interface 3 ($p=0.06$). Another marginally significant difference was also found when comparing the ratings of participants who viewed interface 1 with the ratings of participants who viewed interface 2 ($p=0.057$).

These results indicated that the participants who viewed the two agent interfaces (interface 2 and 3) more strongly believed that the judge was fair than in the single agent interface (interface 1). In the two agent interfaces, agent Jerry was judged by agent David and in the single agent interface Jerry was self-judged. The results point out that the social norm: "independent people's judgement is fairer than self-judgement" is applicable in the human-agent communication scenario.

When comparing the rating of participants who viewed interface 4 with the rating of participants who viewed interface 3, t-test showed significant rating difference ($p=0.043$). Another significant rating difference detected by the t-test was between participants who viewed interface 4 and the participants who viewed interface 2 ($p=0.04$).

The results showed that the participants who observed the interface 2 or interface 3 more strongly believed that the judge was fair than the participants who observed interface 4. Interfaces 2, 3 and 4 all belong to the two agent interfaces, where interfaces 2 and 3 showed

neutral expression or compliment expressions and interface 4 presented opposite expressions. This result demonstrated that the opposite expressions made the participants feel that the judge was against them and was not fair. The opposite expressions presented by the agent strongly decreased the perceived fairness feeling. This result highlighted the influences of emotion expressions presented in the human-agent communications.

No significant difference was found when comparing the mean ratings between interface 3 and interface 2. However, the mean rating for the interface 2 (neutral interface) is higher than interface 3 (complimentary interface). The results indicated that the agent with neutral expression or complimentary expression was perceived as the fairest agents and the neutral expression agent scored highest in the fairness rating. This can provide a guide for human-agent design. In a multi-agent communication system, it is not avoidable for a user to interact with two or more agents. When an agent needs to judge or value a user's answer or ideas, it is best to use a separate agent with neutral expressions or complimentary expressions to present the results.

Question 2: The friendliness of the judge to you

The mean participants' ratings (out of 10) for the question "the friendliness of the judge to you (the participants)" for the different interfaces and the standard deviations are shown in table 45.

	Interface 1	Interface 2	Interface 3	Interface 4
Mean rating	4.18	4.93	7.14	4
Standard deviation	2.5	2.3	1.6	2.0

Table 45: Mean ratings and standard deviations for question 2

It can be seen that interface 3 (the interface with complimentary expressions) achieved the highest mean ratings. This indicated that the participants perceived that the judge in interface 3 was the friendliest. To make a clear view, the mean ratings are visualised in figure 74. It can be seen clearly that there is a strong increase of mean ratings in interface 3.

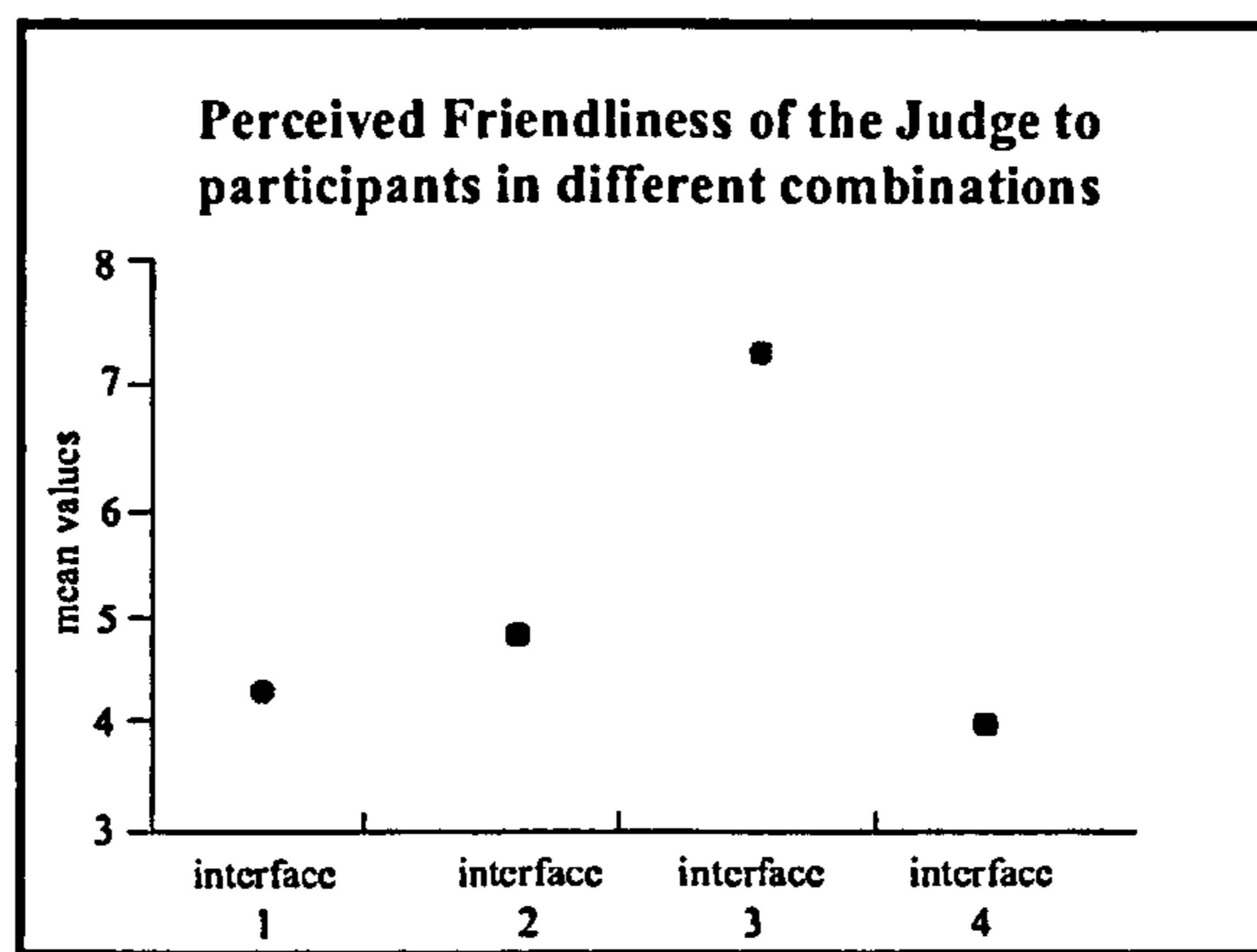


Figure 74: Participants mean ratings of question 2

The t-tests were carried out and marginally significant difference were found when comparing the ratings of participants who viewed interface 2 with the ratings of participants who viewed interface 3 ($p=0.08$). The t-tests also showed that significant difference when comparing the ratings of participants who viewed interface 3 with the ratings of participants who viewed interface 4 ($p=0.02$). Another significant difference was detected when comparing the ratings of participants who viewed interface 3 with the ratings of participants who viewed interface 1 ($p=0.04$).

The significant mean rating difference illustrated that the participants believed that the complimentary expression display agent was more friendly compared to the agent in other interfaces. This demonstrated that emotion expressions can strongly influence the perceived feelings of friendship. The t-tests did not show any significant difference between the ratings of interface 2 (neutral expression) and interface 4 (opposite expression) where $p=0.18$, which mean that there was no significant difference of the perceived friendless between the neutral expression and the opposite expression. This may be explained by two reasons: one reason is that opposite expression may not necessarily create a less friendly person. The other possible reason is that participants may believe that judge should be less friendly or people believe neutral expression is as friendless as opposite expression.

The analysis results demonstrated that an emotional agent can create the friendship feelings with the users. The guides for human-agent design are clear: in a multi-agent communication system, especially in a leisure-oriented environment (e.g., online shopping, online game etc.), an agent with a smile can achieve much more favourable response than a neutral face. This demonstrates the importance of emotion expression in computer agent system.

The result analysis may be extended to the online human-human communication (e.g. a chat room environment). A user's emotion expression display may strongly influence other people's perceived judgements of the user. When chatting online, it should always be remembered that your emotion expression not only reflects what you think but also influences the feelings of others who look at you.

Question 3: The friendliness of the judge to the question agent

The mean participants' ratings (out of 10) for the question "the friendliness of judge to the question agent" for the different interfaces and the standard deviations are shown in table 46.

	Interface 1	Interface 2	Interface 3	Interface 4
Mean rating	Not applicable	4.07	4.71	4.92
Standard deviation	Not applicable	2.5	2.1	2.4

Table 46: Mean ratings and standard deviations for question 3

It can be seen that the participants who viewed interface 4 (interface with opposite expressions) achieved the highest mean rating compared to the other interfaces. To make a clear view, the ratings are visualised in figure 75.

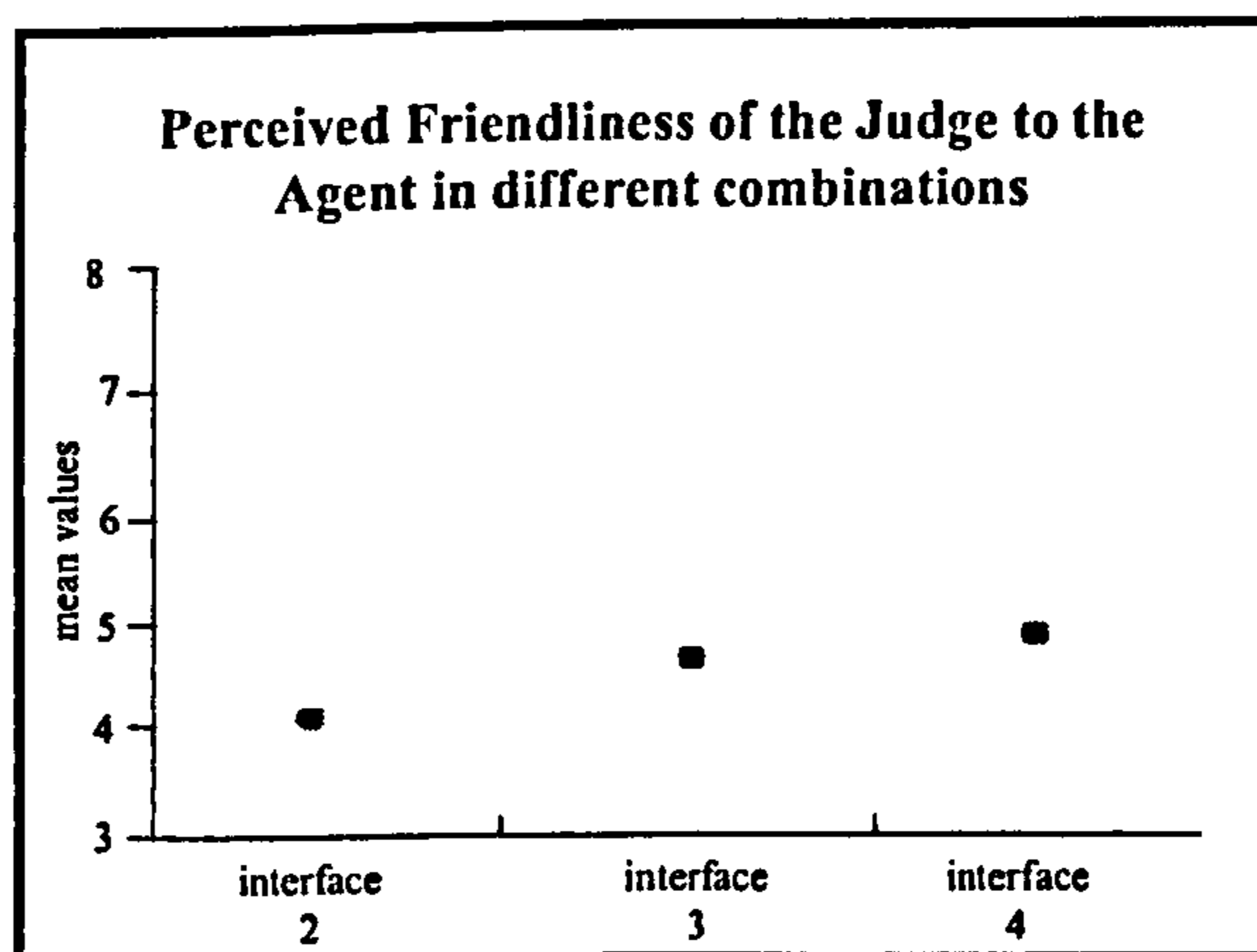


Figure 75: Participants mean ratings of question 3

It can be seen that there are no strong increases this time. The largest increase is from interface 2 to interface 4. The t-tests were carried out and a marginally significant difference was found when comparing the ratings of participants who viewed interface 4 with the ratings of participants who viewed interface 2 ($p=0.087$). From the analysis of question 2, it can be seen that most participants perceived the judge as a friend in interface 3 – the complimentary

expression interface. In other words, in interface 3, the judge was *with* the participants. In interface 4, the judge showed opposite expressions to the participants. Will this mean that the judge was friendlier with the agent this time? The marginal significance illustrated that the judge was friendlier to the agent in interface 4 than in interface 2. However when comparing the mean rating of interface 4 with interface 3, the results were not significantly different. In this case, the feeling "judge David was more friendly to agent Jerry in interface 4 than in other interfaces" could not be supported for all the interfaces.

The mean rating of interface 3 was higher than interface 2, although the t-test showed that they were not significantly different ($p=0.28$). In interface 3, the judge was *with* the participants but there is a higher rating than interface 2, in which the judge was neither *with* nor *against* the participants.

The questions 2 and 3 actually asked the perceived feelings of each participant about the friendliness of the judge to the participant and to the agent. The judge in interface 3 (complimentary expression) was *with* the participants, but it actually did not mean that the judge was definitely *against* the agent. The implication such as "if the judge is with us, then it should be against the agent" was not universal and it was quite weak in the human-agent interaction. Similar ideas may be applied to the analysis of interface 4. The opposite expressions in interface 4 demonstrated that the judge was *against* the participants; however, it actually did not directly mean that the judge was *with* the agent. This showed that emotion expressions can not be transmitted to third parties in the human-computer-agent interaction.

The results demonstrated that the implied emotion or transmitted emotion may not succeed in the human-agent communication. The guides for human-agent design are clear: if an agent wants to present emotional feelings to other agents or users, the agent should show the emotion directly to the object. The transmission of emotion to the third parties is not supported. This shows that the human-agent communication system designers should always keep a clear-cut relationship between agents and human users.

Question 4: Participants' perceived performance of the test

The mean participants' ratings (out of 10) for the question "Your perceived performance of the test" for the different interfaces and the standard deviations are shown in table 47.

	Interface 1	Interface 2	Interface 3	Interface 4
Mean rating	3.36	5.21	6.28	4.91
Standard deviation	1.36	0.83	0.95	1.08

Table 47: Mean ratings and standard deviations for question 4

Interface 3 (interface with complimentary expressions) achieved the highest mean rating which indicates that the participants who viewed interface 3 believed that they performed best. Actually all the participants in the four interfaces performed equally well. To make a clear view of the ratings, the ratings are visualised in figure 76.

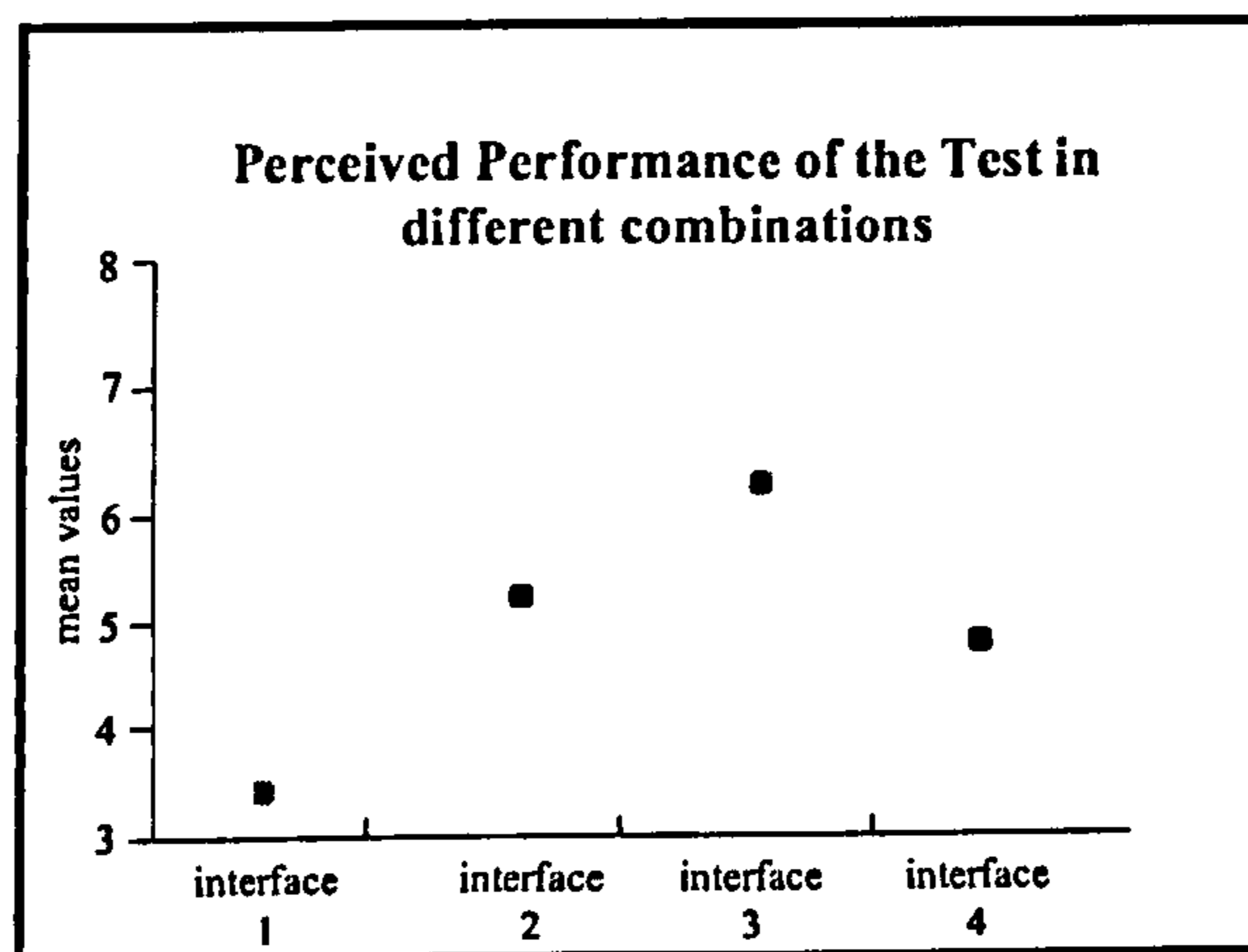


Figure 76: Participants mean ratings of question 4

It can be seen that the mean rating of interface 1 is the lowest. There is a strong increase of mean ratings for all other interfaces compare to interface 1. T-tests were carried out and significant rating differences were found when comparing the ratings of participants who viewed interface 1 with the ratings of participants who viewed other interfaces ($p=0.003$ for interface 2; $p=0.001$ for interface 3; $p=0.003$ for interface 4).

When comparing the ratings of participants who viewed interface 3 with the ratings of participants who viewed other interfaces, significant differences were found ($p=0.001$ for interface 1; $p=0.04$ for interface 2; $p=0.006$ for interface 4). These results demonstrated that the participants who viewed interface 3 believed they performed best and the participants who viewed interface 1 believed they performed the worst even the actual performances for all interfaces were the same. The perceived best performance in interface 3 may be contributed by the independent judgement and the complimentary emotion expression display. As discussed in question 3, the participants who observed the interfaces 2 or interface 3 feel that the judge was

fairer than the participants who viewed other interfaces. With the complimentary expressions displayed by the judge, more participants feel that they carried out the experiment quite well.

The worst perceived performance in interface 1 may be explained by two factors: the first factor was the lack of interaction. The main difference between interface 1 and the other interfaces was the single agent interaction. There were fewer interactions between agents and the participants in interface 1. Fewer interactions may mean less involvement. Less involvement may cause less interest and the perceived performance may be influenced. The second factor was the lack of emotion expressions. Facial expression is vivid, attractive and affective. With facial expression displays, the involvement and confidence may be improved.

It is interesting to re-examine the significant difference between the mean ratings of interface 4 and interface 1. The interface 4 showed the opposite emotion expressions. The opposite expressions should decrease the performance ratings. However the impact of the negative expression was not significant ($p=0.28$ when the mean ratings between interface 4 and interface 2, the neutral expression interface, were compared). This illustrated that negative emotion expression may not be the dominant factor that makes the perceived performance worse. In contrast, the significant mean rating differences between interface 3 and all other interfaces suggested that complimentary emotion expression was the most important factor to improve perceived performance.

As the perceived performances of all the interfaces with two agents were significantly better than the one agent interface, it revealed that the lack of interaction was the most important factor to decrease the perceived performance within different interfaces. In summary, the ratings of question 4 ("Your performance in the test") were influenced by two factors, i.e., interaction and emotion expressions. Complimentary emotion expressions were the main factor in improving the perceived performance and the lack of involvement was the dominant factor for lowering it.

The results demonstrated that complimentary emotion expressions significantly influenced the perceived performance in human-agent interaction. The results illustrated the importance of the emotion expressions in the human-agent communication. The guides for human-agent design are obvious: No one wants to have a poor perceived performance for a communication system. To improve the perceived performances, the complimentary emotion expression on the screen is a key factor. In addition, enough involvement with the user should always be provided.

Question 5, gender and sociability

For question 5 "The easiness of the quiz test", the mean values of the participants' ratings (out of 10) for the different interfaces and the standard deviations are shown in table 48.

	Interface 1	Interface 2	Interface 3	Interface 4
Mean rating	5.18	4.78	5	4.75
Standard deviation	1.7	1.5	1.02	1.9

Table 48: Mean ratings and standard deviations for question 5

The mean ratings of the different interfaces were quite similar and no significant rating difference was found between the four interfaces. The perceived easiness of the test was not found to be influenced by either the emotion expression or the involvement feelings.

In the experiment, every participant's gender and sociability were recorded. The answers for the five questions were grouped according to the gender and sociability and t-tests were carried out to detect whether the mean ratings between different groups were significantly different. However, t-tests did not find any significant differences related to gender or sociability. This may indicate that these two factors do not affect social norms and emotion expression feelings.

10.1.7 Experiment conclusions

An experiment assessing the degree of conformity to social norms and the effect of emotion expressions within human-agent interaction was conducted. The experiment results demonstrated that participants treated computer agents as if they were dealing with humans. Particularly, it was found that participants apply social norms when interacting with computer agents. The assessed social norm was *independent people's judgement is fairer than self-judgement*. The experiment also revealed the functions and the importance of emotion expression for the area of human-agent communication. The results demonstrated that emotion expression can strongly influence the perceived feelings of the interface users.

The results provided numerous practical points for the HCI design of a computer agent system. No matter whether the system is a real-time communication system or an off-line system (e.g. an agent game or learning software), emotion expressions should always be considered as an important factor. It can improve the perceived performance of the user and makes the human-

computer interaction friendlier. Adding other agents may require some additional work but failure to do it may strongly influence the performance of the human-agent interaction.

10.2 Factors that may affect the influence of expressive images

The previous section demonstrated that presenting expressive images will significantly influence user's feelings. In this section, an experiment assessed the factors that may influence the perceived emotional feelings are discussed. The experiment investigated two factors – the expressive image intensity and the display duration (the wear down effect).

10.2.1 Background

Evatt [1997] discovered that the perceived salience of a public policy issue will increase when news about the issue is presented in a highly emotion-evoking manner and decrease when the news about the same issue is presented in a less emotion-evoking manner. This demonstrates that when the intensity of the emotion-evoking manner increases, the salience the readers perceived will increase.

Hovland et al. [1953] suggested that increasing the intensity of the emotion-evoking content might not always elevate salience. At a certain intensity level, the effect could start to drop off. Hughes [1992] and Kinnick [Kinnick et al. 1996] demonstrated a wear-down phenomenon. Participants' favourable responses would be reduced after the emotion-evoking manner was repeated over a long period of time. A typical wear down effect can be described as such: when an individual first meets an exciting stimulus, the excited feelings will be high. When stimulus is repeated many times, the exciting feelings will not continue to rise, instead the feelings will be stable or even fall if the stimulus is endless. The problem with assessing wear-down factors is that it is hard to predict the exact time that feelings will be stable or will fall, and Evatt [1997] demonstrated that the wear-down phenomenon is not always observed. The wear down factor is illustrated visually in figure 77.

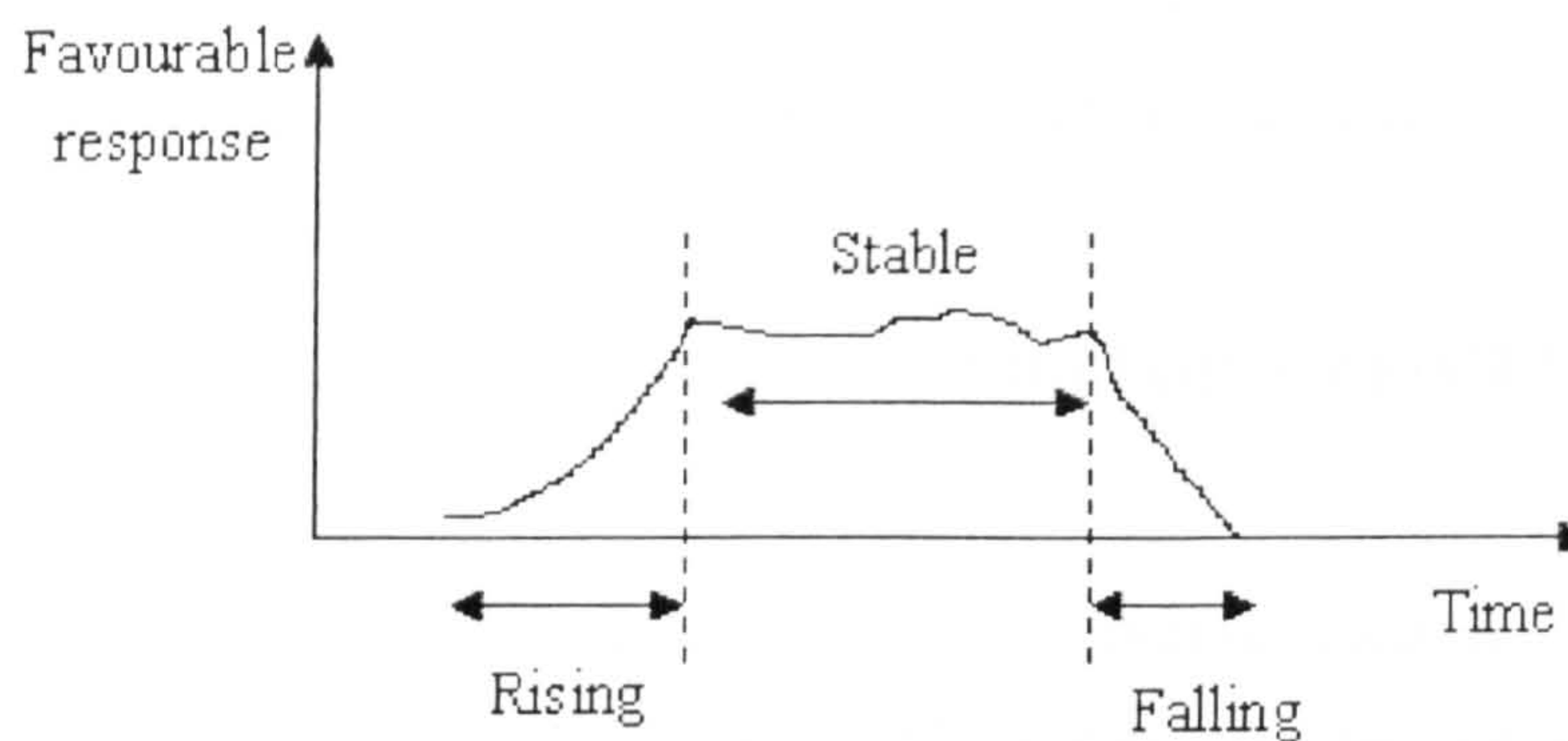


Figure 77: The wear down factor

It can be seen that the effects of presenting textual information in an emotion-evoking manner will be influenced by different factors. However above results were purely based on the study of textual messages, which means that the emotion-evoking manners used were pure text. Will the emotions presented by expressive images produce the same phenomena?

10.2.2 Planning of experiment

An experiment was designed to assess the factors that may influence the expressive image display. A test platform was developed using flash, JSP and HTML programming. The platform included an online presentation session and a questionnaire session. Three hypotheses were made to identify the expected experiment results.

- **Hypothesis 1:** When the intensity of the expressive images increases, the perceived emotional feelings will increase. This hypothesis aimed to demonstrate that by increasing the intensity of the emotion-evoking manner - expressive images, the perceived feelings would increase although the accompanied text remained the same.
- **Hypothesis 2:** When the intensity of the expressive images raises beyond a realistic level, the perceived feelings will stop increasing and start to decrease. This hypothesis aimed to demonstrate the ceiling phenomena. The levels of feelings were predicted to fall as the participants were exposed to extremely high intensity expressive images.
- **Hypothesis 3:** After viewing three stories, the perceived feelings for the third story will be higher for people who viewed a story accompanied with medium intensity expressive images following two stories accompanied with neutral expressive images than for people who viewed the same stories each accompanied with medium intensity expressive images.

10.2.3 Methods of the experiment

10.2.3.1 The emotion intensity test

To assess the influences of expressive images with different intensities, a human-like software agent was developed. The agent presented a story on the screen and offered facial expression display to experiment participants. To focus on the influences of expressive facial images, the story itself contained minimal emotional content.

A between groups experimental design was applied for this experiment. In all conditions the agent presented the same stories to every participant, however in the first condition (low intensity condition), the agent presented facial images with low expression intensity to the participants. In the second condition (medium intensity condition), the agent presented facial images with medium expression intensity and in the third condition the agent presented extreme expression intensity to participants (extreme intensity condition). Typical screens of all conditions are shown in figure 78.

After viewing the story presentation session, all participants in the three groups were directed to the same questionnaire session. The applied questionnaire is based on the "Personal Involvement Inventory" (PII) developed by Zaichkowsky [1987] and the questions are listed in table 49.

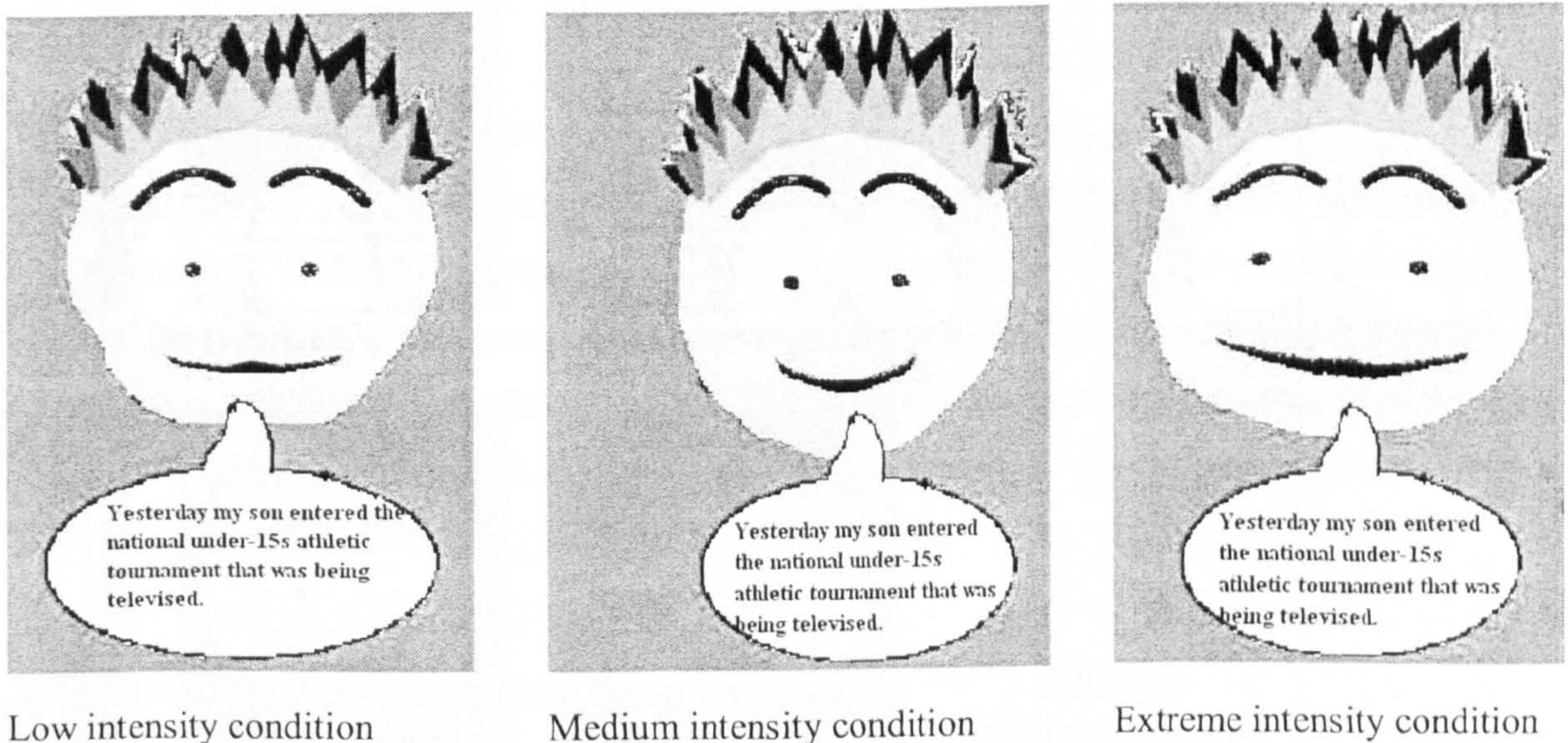


Figure 78: Typical Screens of three conditions

Question 1	How important the story is?
Question 2	How interesting the story is?
Question 3	How exciting the story is?
Question 4	How fascinating the story is?
Question 5	How neutral the story is?
Question 6	How much do you feel involved in the story?

Table 49: Questions asked in the questionnaire session

10.2.3.2 Display duration - wear down effect test

To assess the wear down factor, a similar test method as the emotion intensity test was used again. However, instead of presenting only one story, the agent presented three stories, all of which contained minimal emotional content in order to keep the focus on the expressive images. The presentations to the different groups differed only in the intensity of the expressive facial images. In the first group, the agent presented two stories with neutral facial expressions followed by presenting a story with medium intensity facial expressions. In the second group, the agent presented all the three stories with medium intensity expressions. The typical screens of group 1 and group 2 are shown in figure 79.

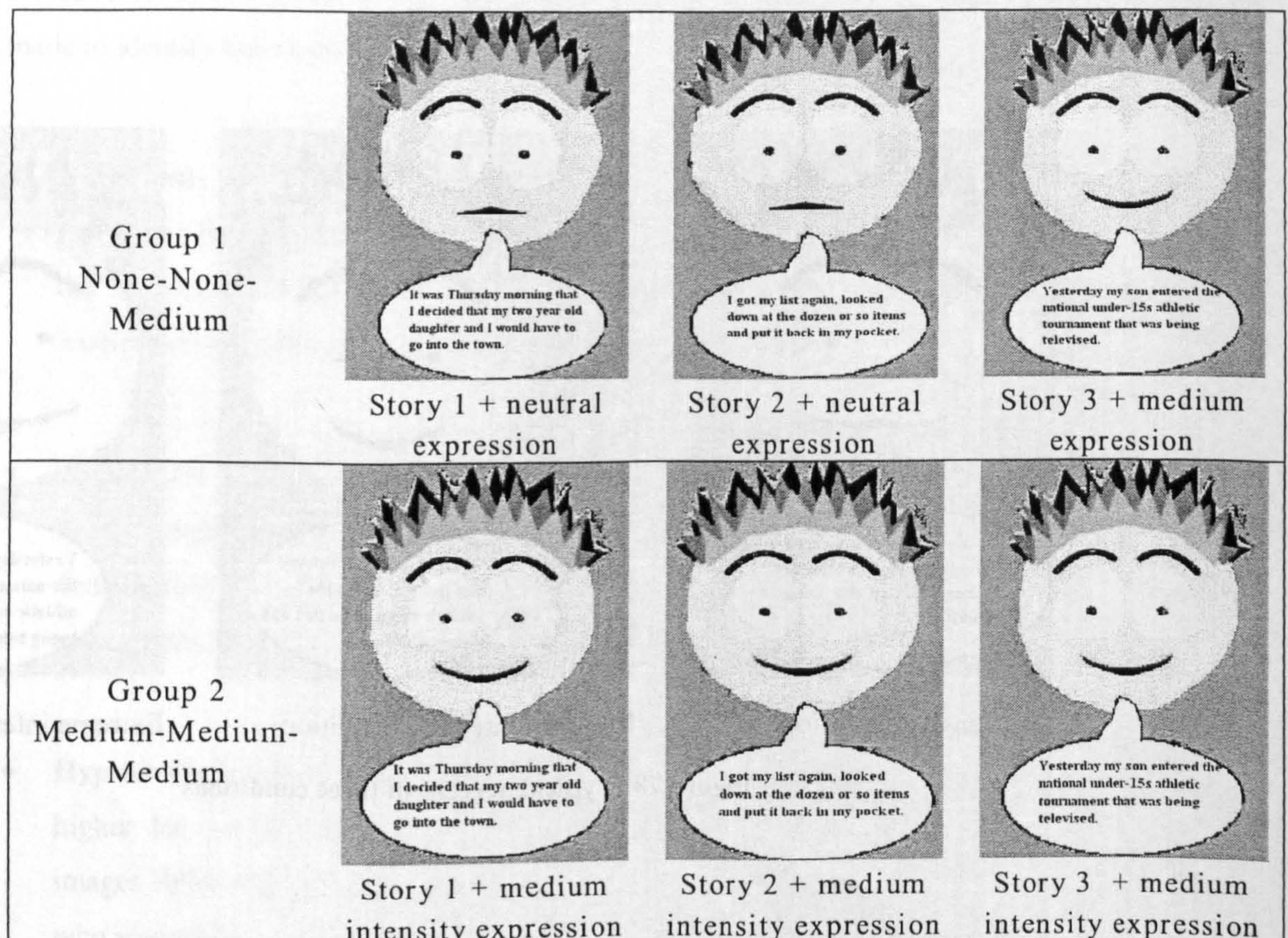


Figure 79: Typical screens of the two groups

After viewing the story presentation session, all participants were directed to the same questionnaire session. The applied questionnaire was the same as the above questionnaire used in the intensity test, which has been shown in table 49. The hypothesis 3 predicted that the perceived emotional feelings would be higher when participants viewed medium expressive images after two sets of neutral expressive images. the test is only concerned with the responses to the third story in each group. The design for each group of story presentations is shown in table 50.

Participants	Expressive image intensity		
	Story 1	Story 2	Story 3
Group1	None	None	Medium
Group2	Medium	Medium	Medium

Table 50: Description of the experiment design to test hypothesis 3

10.2.4 The experiment operation

10.2.4.1 The intensity test operation

Sixty students and staff from Bournemouth University participated this online experiment through the Internet. The experiment included two sessions. First a cartoon agent presented a story to each participant and then participants answered a questionnaire about the emotional feelings they perceived.

10.2.4.2 The wear down test operation

Forty students and staff from Bournemouth University participated in this online experiment. The participants performed the experiment through the Internet as well. The participants were divided into two groups. One group viewed the none-none-middle intensity setting and the other group viewed the middle-middle-middle intensity setting. The cartoon agent presented three stories to each participant and then the participants answered a questionnaire about the perceived emotional feelings.

10.2.5 How the data was measured and analysed

For the intensity test, the main measurements were the answers of the questionnaire, the values vary from 1 to 10, which 1 means lowest, 10 means highest. The preference of interfaces was examined by testing the hypotheses described above. For the wear-down test, although participants viewed three interfaces, only the answers in the final questionnaire were collected and the first two stories were to create the atmosphere. As gender and participant' sociability did

not show significant influence in the previous experiment, these two factors were not recorded in these two tests.

10.2.6 Experiment result analysis

10.2.6.1 Intensity test results

For the low intensity condition, the mean perceived emotion feeling is 3.22. In the medium intensity condition the mean perceived emotion feeling is 4.6. For the extreme intensity condition, the mean perceived emotion feeling is 3.7. T-test was carried out to compare the rating differences between the low intensity condition and medium condition and the results revealed a significant difference ($p=0.044$). Therefore it is possible to accept hypothesis 1 that when the intensity of the expressive images increases the perceived emotional feelings will increase. The t-test showed a marginally significant difference between the ratings of the medium condition and the extreme high condition ($p=0.08$). The result indicates that hypothesis 2 is correct on asserting that when the intensity of the expression images rises beyond a realistic level the perceived feelings will stop increasing and may fall.

10.2.6.2 The wear down result analysis

For the none-none-medium condition, the mean value of the perceived emotional feeling for story 3 was 4.4. In the medium-medium-medium condition, the mean perceived emotion feeling was 4.7. The t-test revealed no significant difference between the ratings of the two groups. Thus hypothesis 3 was not supported by the test result.

10.2.7 Experiment results discussion

The experiment results supported hypothesis 1 and 2, which predicted that the perceived emotional feelings from textual stories are strengthened when the story is accompanied with suitable expressive images. It was first hypothesised that when the agent presents a story with suitable expressions, the perceived emotional feelings will increase. As predicted, the participants who read the story with medium expressive intensity images perceive more emotional feelings than the participants who read the story with low intensity images. A summary of the experiment results is presented in table 51.

Hypothesis	Description of hypothesis	Test results
1	The perceived emotional feelings will increase when the intensity of the expressive images increases.	True
2	The perceived feelings will fall when the intensity of the expressive image rises beyond a realistic level.	True
3	Wear-down effect is valid in Internet communication	False

Table 51: Results of the tests of hypotheses

The next question is whether a ceiling exists. Will the gain achieved from increasing expression intensity be lost when the intensity reaches an unrealistic level? It was predicted that the extremely high intensity expressive images may decrease the perceived emotional feelings. The experiment results partially supported this hypothesis as a marginally significant difference between the two conditions was found.

Participants reported a significantly higher level of emotional feelings in the medium intensity condition than in the low intensity condition. When the expressive facial images were exaggerated to an unrealistic level, the reported perceived emotional feelings decreased.

Hypothesis 3 states that the influence of external factors, such as the wear-down effect, would affect the perceived emotional feelings. However the results show that the perceived emotional feelings remain stable. This may suggest that the perceived emotional feelings are independent of external factors in particular the number of times the expressive images are displayed. The effect that an expressive image has on the viewer is not changed whether the viewer has seen the image many times or whether it is the first time it has been displayed. Another explanation is that the compared data are both within the stable phase, and the display should be kept longer to move to the down phase.

The experiment results indicated that emotion intensity affect the perceived feelings. It also indicated that external factors such as the wear-down effect may not influence the perception. Some practical implications can be drawn for the design of affective computer interactions. First, to convey emotional feelings, expressive images should be provided during human-computer interaction. Second, medium intensity expressive images can achieve the best performance. Both decreasing the intensity and increasing the intensity to an extreme level will show negative influences to the perceived feelings. Third, the perceived emotional feelings may be independent of the factors such as display duration or the stable phase is quite long. This means

that the expressive images can be shown as soon as appropriate and there is no need to worry that the perceived emotional feelings may decrease with reasonable repeated use.

10.2.8 Experiment conclusions

An experiment was conducted to test the factors that may influence the perceived emotional feelings when users interact with emotional agents. The experiment results demonstrated that the internal factor such as the intensity of expressive images did significantly influence the perceived feelings. The perceived emotional feelings did increase when the intensity of expressive image increases. However, when the intensity of expressive images increased to an unrealistic level, the perceived emotional feelings decreased.

The external factor - display time (wear down effect) was found not to be significantly influence the perceived feelings. The research indicated that expressive images did influence the perceived emotional feelings and as long as the display was valid, the appropriate expressive images could be shown and there was no decrease of perceived emotional feelings due to a wear-down effect.

10.3 Chapter conclusion

This chapter presents two experiments, one to assess the influences of display expressive images and the other to assess the factors that may affect the influences. An expressive agent interface was created as the test platform for the first experiment and the experiment results demonstrated that expressive images do influence users' perceptions. Although only tested in a human-agent interaction context, the detected influences may also be valid for other environments (e.g. online chat and online game interfaces). The second experiment examined two factors that may affect the influences of expressive images and the experiment results supported the first two hypotheses. This chapter presents the last issue examined by this research project. The next chapter will give the conclusions of this research work and present research suggestions for the future.

Chapter 11

Conclusions and further research suggestions

The aims of this research work were to investigate the methods of detecting and visualising emotional feelings for Internet communications and identify the influences of automatic emotion visualisation on Internet. The study primarily addressed the issues of emotion detection from text and expressive image synthesis, and focused on the identification of the influences of automatic emotion detection and visualisation. The main findings of this research work are within the context of Internet communication; however, the findings may be relevant to emotion detection and visualisation in other channels such as audio or video communications over Internet.

The first phase of the project focused on the methods of detecting and extracting emotional feelings from textual messages. An emotion momentum theory was developed and then an emotion extraction engine based on this theory was implemented. A series of experiments were conducted to assess the performance of the emotion extraction engine and the experiment results demonstrated that the emotion extraction engine is fast and accurate enough to analyse sentences collected in real-time communication environment.

The second phase of the project was to synthesise facial expressive images that can be used by the emotion extraction engine to represent the emotional feelings. An expressive image generator was implemented based on image warping and morphing approaches. The expressive image generator can synthesise expressive images belonging to six expression categories: happiness, sadness, surprise, disgust, fear and anger. For each category, three different intensities – low, middle and high were achieved. Although the emotion extraction engine only requires discrete facial expression images, a facial expression animation feature is provided by the expressive image generator as well. A series of experiments were conducted to assess the performance of the expressive image generator and the results showed that the generated facial expressions were successfully recognisable to most participants in the experiments.

With the emotion extraction engine and the expressive image generator, an emotion visualisation framework prototype was set up. In the final phase of the project, a series of experiments that aimed to assess the influences and other relevant factors of presenting

visualised emotion to Internet users were carried out. The experiment results confirmed the hypothesis that users prefer the interfaces with emotion visualisation feature and visualised emotion does influence perceived feelings.

A series of recommendations on how to automate emotion communication into an Internet interface can be derived from the development work. A series of suggestions of the influences of emotion on Internet interfaces can be deduced from the results of the experiments.

The original contributions to knowledge of this study are the emotion momentum theory, the innovations built into the prototype emotion extraction engine, the fast and user friendly interface of the expressive image generator and the identification of the influences of emotion visualisation to online interfaces.

11.1 Conclusions and further research suggestions for the emotion extraction engine prototype

The emotion momentum theory was proposed as the fundamental theory to support the analysis of the emotional feelings embedded in textual messages of Internet communication. The emotion momentum theory proposes the following statements:

1. People stay in certain emotional states and resist changes in their emotional states.
2. The value of the emotional momentum is proportional to the intensity of the emotional state.
3. Emotional momentum is affected by interaction with people. The momentum has a sign, and can be positive, negative or neutral.
4. Personal emotions eventually die out or 'decay' in time but the rate of decay is affected by positive or negative interactions with other people and sharing of emotions.

The emotion extraction engine prototype implemented the emotion momentum theory and enabled a range of features including easy to use interface, visualised emotion display and voice generated by the Microsoft speech engine. The main phases of development of the prototype included the implementation of the tagging system, the sentence analysis (sentence parsing) system integrating real-world knowledge analysis, the fuzzy logic components and an easy-to-use interface. The features that make the emotion extraction engine user friendly include the

clear and consistent uncluttered layout of the screens and the use of text, text-to-speech and facial expressions to present information.

The emotion extraction engine includes three components: tagging system, parser and the interface layer. The tagging system contains a tag set and an 18,000 word dictionary. The emotional words contained in the dictionary are tagged to represent the corresponding emotion categories and intensities. Totally six emotion categories *happiness, sadness, disgust, fear, surprise* and *anger* are classified in the tagging system. The intensities of emotional words for every category are defined as low, middle or high. The parser component implements not only linguistic analysis, but also real-world knowledge analysis and fuzzy logic components. To analyse mood and conflicting emotions, mood average methods and fuzzy logic analysis were implemented as well.

Two performance assessment experiments were carried out. The first experiment assessed the speed of a 2D chat interface, in which the emotion extraction engine was embedded. The experiment results demonstrated that the engine's analysis is not a speed bottleneck, the chat server's speed is not a bottleneck and the data transmission is the most time-consuming part. The speed experiment results illustrated that the engine was fast enough to be applied in real-time Internet. The second experiment was to assess the accuracy of the emotion extraction engine. Two types of texts, i.e., text messages collected from chat archives and sentences collected from published articles, were input into the emotion extraction engine. The experiment results demonstrated around 87% sentences in the collected chat archive were correctly analysed and around 76% sentences in the published articles were correctly identified by the emotion extraction engine. The results showed that the accuracy of the extracted emotion was acceptable for Internet communication purposes.

The emotion extraction engine was developed to a sufficient level in order to demonstrate aspects of emotion detection and extraction. However, some individual components of the system could be further developed and tested, such as implementing alternative tagging methods (e.g. statistical methods).

The tagged dictionary was developed by a group of researchers. However, it would be interesting to apply standard dictionary such as the Brown corpus and BNC or have two groups of researchers to develop the dictionary independently and compare the possible performance differences.

The development work of the emotion extraction engine and the experiments showed that by limiting the context the engine is applied into, the emotion analysis could achieve acceptable performance. It would be interesting to design a context aware framework or context aware language to describe different contexts.

The emotion momentum theory assumes that every user starts with a neutral emotion and the changing pace is same for everyone. In the future, the emotion momentum theory may be extended to comprise user specific factors, such as personality, background, or even religion to create a more specific and tailor made emotion model.

11.2 Conclusions and further research suggestions for the expressive image generator

The expressive image generator was designed to creating facial images representing the emotions sensed by the emotion extraction engine. The users of the expressive image generator were not expected to be expert computer users, so an easy-to-use interface is required. An effort was made to provide a user friendly prototype expressive image generator.

The main phases of development of the prototype included the identifications of distinctive expressive features for human and cartoon faces, the implementation of an image warping subsystem, the implementation of an image morphing subsystem and the implementation of facial animation features. The expressive image generator includes two components: image warping subsystem and the image morphing subsystem. The image warping subsystem implements local area warping and the morphing subsystem applies the back morphing technologies. To provide an easy-to-use user interface, the complexity of the user interaction is restricted to the minimum.

Two performance assessment experiments were carried out. One experiment tested the synthesised expressive human facial images and the other experiment assessed the generated expressive cartoon facial images. Both experiments measured the recognition rates of the expressive images in two interfaces, i.e., an interface presented only the expressive images and an interface presented the expressive images with suggestive texts. The cartoon experiment results demonstrated that more than 59% of the participants correctly guessed the expressive images in image-only interface and more than 78% of the participants recognised the expressive images in the image plus text interface correctly. For human facial images, more than 64% of

the participants correctly recognised the images in the image only interface and more than 76% of the participants recognised the expressive images in the image plus text interface in the middle and high intensity.

The expressive image generator was successfully developed to a sufficient level in order to demonstrate aspects of facial expression synthesis. However, it might be worthwhile to generate features such as alternative facial expression categories (e.g. the categories defined in OCC model) and study different image manipulation algorithms to improve the performance of the expressive image generator, especially the recognition rate.

The expressive image generator can only synthesise 2D facial expressions. More work can be done in synthesising 3D facial expressions for real-time Internet communication. Similar ideas and techniques used in the 2D facial expression synthesis may be reused in the 3D field and new approaches may need to be proposed.

The corresponding experiments assessed the recognition rate of the generated images. It would be interesting to examine the acceptance rate of the synthesised images as “accept” and “recognise” may lead to different meanings (e.g. I recognise a happy expression from this image, but may not think that expression is an acceptable happy expression). The testing methods would be similar to the experiments described in chapter 6 that assessed the recognition rate except that the used questionnaire should be redesigned.

11.3 Conclusions and research suggestions for embedding emotion extraction engine into different applications

To demonstrate the variety usage of the emotion extraction engine, several applications have been developed. A 2D expressive interface and a 3D expressive interface were created to demonstrate how to directly embed the emotion extraction engine into different environments by simply changing the interface layer. It is worthwhile to investigate more areas the emotion extraction engine and the expressive image generator can be applied into.

The development of the stock market emotion analyser demonstrated how to revise the fundamental components of the emotion extraction engine, i.e., the tagset and tagged dictionary, to fit into stock market context. By applying similar procedures, the emotion extraction engine

may be adapted for various environments.

The stock market emotion analyser itself revealed some interesting phenomena. The experiment result showed that about 70% of articles correctly predicted the next day's movement trend. The experiment also showed that a significant relationship between emotions detected in stock market articles and the next day's overall value exists in the FTSE 100 and FTSE all-index level. From the experiment results, it can be seen that emotion contents do correlate with the stock market. It would be worthwhile to analyse any correlation further between emotion articles and the share movements or the trading volumes of specific business sectors (e.g. telecom or oil industry etc.). In this research, only chose the articles that discussed the general level stock market movement were chosen. Further research may choose articles focusing on individual company or business sectors to detect the possible correlation.

11.4 Conclusions and further research suggestions for experiments assessing the influence of visualised emotions

11.4.1 The interface preference experiment

The emotion extraction engine and the expressive image generator created an expressive communication interface for Internet communication which can detect and visualise emotions embedded in textual messages. Will users prefer the expressive communication interface? The preference experiment was conducted to answer this question. Four platforms were presented to experiment participants and the participants were free to choose the platform and the component they preferred.

The experiment results demonstrated that firstly, most participants (78%) preferred the platform with expressive images, voice and text. A significant number of the participants preferred the platform with expressive images, voice and text much more than the text only platform. Theoretically, an individual's cognitive style influences their preference of the ways knowledge is presented, but no significant influences of cognitive style were found in the preference of the interfaces.

Further research may focus on the assessment of the preference of different interfaces which have more components, e.g. video clips, non-synthesised voice or artistic drawings etc. It would also be interesting to assess the possible preference differences between 2D and 3D facial images. To examine the preference difference between different cognitive styles, participants

were divided into groups according to their cognitive styles. Although no significance was detected, it may be worth grouping people by other factors, e.g., gender, age, sociability and personality etc, as in theory these factors may influence a person's perceptions.

11.4.2 The experiment comparing the preference of the 2D and 3D interfaces

An experiment was set up to assess preference differences between the 2D and 3D expressive interfaces. To examine the different perceived feelings, twelve activities that can be carried out on the Internet were listed, and for each activity the participants were asked to choose their preferred interface. The activities were classified into two styles: business-oriented and social-oriented. The experiment results showed that significant differences exist in the preference of interfaces for different activities. Business-oriented activities were preferred to be carried out in the 2D expressive interface and social-oriented activities were preferred in the 3D expressive interface. Individual participant's sociability was found out to have limited impact on the preference of the expressive interfaces, but the activity style has stronger influences on the preference of 2D and 3D interfaces.

The activities were classified into two general categories. It would be interesting to divide the activities into detailed categories and analyse the interface preference difference. The interfaces compared in this research were a 2D expressive interface and a 3D expressive interface. Further research may be carried out to compare a 2D expressive interface with a 2D neutral interface, a 3D expressive interface with a 2D neutral interface or a 2D expressive interface with a 3D neutral interface to create a clear picture of the interface preferences.

11.4.3 The experiment assessing the influences of emotion visualisation

An experiment aimed to detect the influences of the visualised emotion was designed. Experiment participants were asked to view an online quiz session and an online questionnaire session. In the online quiz, an agent presented nine questions and either the agent itself or another independent agent judged the answers of the participants. When judging the answers, the judging agent either present compliment expressions, opposite expressions or neutral expressions to the experiment participants. The online quiz session was to create an atmosphere and the data was collected in the online questionnaire session. The questionnaire session asked five questions relating to the fairness of the agent and the perceived performance of the participants in the experiment. The results of the questionnaire session demonstrated that participants behaved as if they were dealing with humans. Particularly, participants applied the social norm – "independent people's judgement is fairer than self-judgement" when interacting

with computer agents. The perceived performance of the participants using the compliment expression interface was significantly higher than the participants in the neutral expression or opposite expression interface. The experiment results demonstrated that emotion expression can strongly influence the perceived feelings.

More social norms may need to be verified in order to draw a clear picture of the social factors of the cyber world. This research work identified the influence of visualised emotion upon an online quiz environment. It would be interesting to apply similar ideas into online chat, online game or online shopping environments to detect whether the established conclusions are still valid.

11.4.4 The experiment assessing the factors that may affect perceived emotional feelings

This experiment assessed two factors that may influence perceived emotional feelings. The two factors are display duration (wear-down effect) and expressive image intensity. The experiment results demonstrated that the intensity of expressive images did significantly influence the perceived feelings. The perceived emotional feelings did increase when the intensity of the expressive image increases. However when the intensity of expressive images increased to an unrealistic level, the perceived emotional feelings decreased. The display time did not produce a significant influence on the perceived feelings in the experiment. It thus showed that the wear-down effect did not influence the perceived emotional feelings significantly in this experiment.

The experiments were performed by enough participants to produce reliable statistical data to assess the performance, compare the preference and examine the factors. However more participants are required to perform cross group experiments for complex tasks. Also new experiments needed to be devised in order to fully assess the influences of emotion visualisation.

More research is required to verify the factors that may influence the perception of emotion. In this research work, two factors were examined. However other factors, such as context and mood, etc., need to be examined to create the whole picture of the influencing factors. The background studies on social factors need to be carried out in order to set up a solid ground for further research work.

References

- Adamatti, D. F and Bazzan, A. L. C., 2002. A framework for simulation of agents with emotions. *In: WORKCOMP (Workshop de Computacao), ITA, Sao Jose dos Campos, Sao Paulo.*
- Andre, E., Klesen, M., Gebhard, P., Allen, S. and Rist, T., 1999. Integrating models of personality and emotions into lifelike characters, *In: A. Paiva and C. Martinho, eds. Proceedings of the workshop on affect in interactions - towards a new generation of interfaces in conjunction with the 3rd annual conference, Siena, Italy.* 136-149.
- Arnott, J. L., Alm, N., and Murray, I.R., 1993. Enhancing communication prosthesis with vocal emotion effects, *In: Proceedings of the ESCA tutorial and research workshop, speech and language technology for disabled Persons, Stockholm, Sweden.* 165-168.
- Badler, N. I., Metaxas, D., Webber, B. and Steedman, M., 1995. The center for human modelling and simulation. *Presence*, 4 (1), 81-96.
- Badler, N. I., Bindiganavale, R., Allbeck, J., Schuler, W., Zhao, L. and Palmer, M., 2000. Parameterized action representation for virtual human agents. *In: J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, eds. Embodied Conversational Agents.* The MIT Press, Cambridge, MA, 256-284.
- Badre, A. N., 2002. *Shaping Web Usability: Interaction Design in Context.* Addison-Wesley.
- Bailenson, J. N., Beall, A. C., Blascovich, J., Weisbuch, M., and Raimundo, R. (2001) Intelligent agents who wear your face: Users' reactions to the virtual self. *Lecture Notes in Artificial Intelligence*, 86-99.
- Bailenson, J. N., Blascovich, J., Beall, A. C. and Loomis, J. M., 2001. Equilibrium theory revisited: mutual gaze and personal space in virtual environments. *Presence: teleoperators and virtual environments*, 583-598.
- Bartneck, C., 2002. Integrating the OCC model of emotions in embodied characters. *Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges, Melbourne.*
- Bates, J., 1992. Virtual reality, art, and entertainment. *Presence: The journal of teleoperators and virtual Environments*, 133-138.
- Bates, J., 1994. The role of emotion in believable agents. *Communications of the ACM*, 37 (7), 122-125.

- Beard, S., Crossman, B., Cechner, P. and Marriott, A., 1999. FAQBot. *In: Proceeding of pan Sydney area workshop on visual information processing, Sydney, Australia.*
- Bellamy, A. and Hanewicz, C., 1999. Social psychological dimensions of electronic communication. *Journal of Sociology* [online]. Available from: <http://www.sociology.org/content/vol004.001/bellamy.html> [Accessed 20 May 2003].
- Bergen, J. V., 2003. *All you need to know about stocks* [online]. Available from <Http://stocks.about.com/> accessed 16/May/2003 [Accessed 12 April 2003].
- Bermond, B., Nieuwenhuyse, B., Fasotti, L., and Schucerman, J., 1991. Spinal cord lesions, peripheral feedback, and intensities of emotional feelings. *Cognition and Emotion*, 5, 201-220.
- Bezdek, J.C., 1989. The Coming Age of Fuzzy Logic. *In: J. C. Bezdek, ed. Proceedings of IFSA Congress.*
- Brehm, S., Kassin, S. and Fein, S., 1999. *Social psychology*. 4nd ed. NY: Houghton Mifflin.
- Brousseau, E., 1990. Information technologies and inter-firm relationships: The spread of inter-organization telematics and its impact on economic structures. *In: 8TH international conference of International Telecommunications Society.*
- Bryson, S. and Levit, C., 1991. The virtual wind tunnel: an environment for the exploration of three-dimensional unsteady Fluid Flows. *In: Proceedings of IEEE Visualisation '91, San Diego, Ca.*
- Bulthoff, H. H. and Edelman, S., 1992. Psychophysical support for 2D view interpolation theory of object recognition. *In: Proceedings of the national academy of sciences*, 89 (1). 60-64.
- Burns, T. and Stalker, G. M., 1961. *The management of innovation*. London: Tavistock.
- Camurri, A., Coglio, A., Coletta, P. and Massucco, C., 1997. An architecture for multimodal environment agents. *In: Proceeding of Italian Association for Musical Informatics (IAMI) international workshop.* 48-53
- Cannon, W. B., 1927. The James-Lange theory of emotion: a critical examination and an alternative theory, *American journal of psychology*, 39, 10-124.
- Cannon, W. B. 1932. *The wisdom of the body*. 2nd ed. New York: Norton Pubs.
- Cassell, J., 2000. More than just another pretty face: Embodied conversational interface agents. *Communications of the ACM*, 43 (4), 70-78.

- Cassell, J., Sullivan, J., Prevost, J., and Churchill, E, 2000. *Embodied Conversational Agents*. MA: The MIT Press.
- Cato, J., 2001. *User-centered web design*. UK: Addison-Wesley.
- Chen, L. S., Tao, H., Huang, T. S., Miyasato, T. and Nakatsu, R., 1998. Emotion recognition from audiovisual information, *In: Proceeding of IEEE workshop on multimedia signal processing*, 83 - 88.
- Conati, C., 2002. Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16, 555 - 575.
- Cootes, T., Edwards, G. and Taylor, C., 1999. Comparing active shape models with active appearance models, *In: Proceeding of British machine vision conference*, 173-182.
- Crystal, D., 1996. *Rediscover Grammar*, London: Longman.
- Davis, J .W and Bobick, A., 1998. Virtual PAT: a virtual personal aerobics trainer, *In: Proceeding of workshop on perceptual user interface, San Francisco*. 13-18.
- Dellaert, F., Polzin, T. h. and Waibel, A., 1996. Recognizing emotions in speech, *In: Proceeding of ICSLP 96, Philadelphia, PA*.
- Dix, A., Finlay, J., Abowd, G. and Beale, R., 1998. *Human computer interaction* 2nd ed. Europe: Prentice Hall.
- Dennett, D. C., 1988. Précis of the intentional stance, *behavioral and brain Sciences*, 11, 495-546.
- Driankov, D., Hellendoorn, H., and Reinfrank, M., 1993. *An Introduction to Fuzzy Control*. New York: Springer-Verlag.
- Eberts, R. E., 1994. *User interface design*, Europe: Prentice Hall.
- Edwards, A. D. N., 1995. Computers and people with disabilities. *Extra-Ordinary Human-Computer Interaction, Cambridge Series on Human-Computer Interaction*, 7, Cambridge University Press, 19 - 43.
- Egges, A, Kshirsagar, S. and Magnenat-Thalmann, N., 2004. Generic Personality and Emotion Simulation for Conversational Agents. *Computer Animation and Virtual Worlds*. 15 (1), 1-13.
- Ekman, P., 1996. Why don't we Catch Liars? *Social Research*, 63, 801-818.
- Ekman, P. and Friesen, W. V., 1971. Constants across culture in the face and emotion. *Journal of personality and social psychology*, 17, 124-129.
- Ekman, P., Friesen, W. V. and Ellsworth, P. 1972. *Emotion in the human face: guidelines for research and an integration of findings*. New York: Pergamon Press.

- Ekman, P. and Friesen, W. V., 1978. *Facial Action Coding System: a technique for the measurement of facial movement*. CA: Consulting Psychologists Press.
- Ekman, P. and Oster, H., 1979. Facial expressions of emotion. *Annual Review of Psychology*, 30, 527 - 554.
- Ekman, P., 1999. Facial Expressions. In: T. Dalgleish and M. Power, eds. *Handbook of Cognition and Emotion*, New York: John Wiley and Sons Ltd.
- Ekman, P., Friesen, W. V. and Hager, J. C., 2002. *The facial action coding system*. Salt Lake City: Research Nexus eBook.
- Eldelman, S. and Bulthoff, H., 1992. Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, 32, 2385 - 2400.
- ELSTER, J., 1996. Rationality and the emotions. *The economic journal*, 106 (438), 1386 - 1397.
- ELSTER, J., 1999. *Alchemies of the mind: rationality and the emotions*. UK: Cambridge University Press.
- Elliot, C., 1992. *The Affective Reasoner: A Process model of emotions in a multi-agent system*, Institute for the Learning Sciences. Thesis (PhD). Northwestern University.
- El-Nasr, M. S., Yen, J., and Ioerger, T., 2000. FLAME - A fuzzy logic adaptive model of emotions. *Autonomous agents and multi-agent systems*, 3 (3), 219 - 257.
- Evatt, L. D., 1997. *The influence of Emotion-Evoking content of news on issue salience*. Thesis (PhD). University of Texas at Austin.
- Ezzat, T. and Poggio, T., 1998. MikeTalk: a talking facial display based on morphing visemes. *Computer animation*. Philadelphia, PA, 96 - 102.
- Faceworks 1998. *DIGITAL FaceWorks Animation Creation Guide* [online]. Available from: <http://interface.digital.com/download/Face Works.pdf> [Accessed 03 July 2003].
- Fernandez, R. and Picard, R. W., 1997. Signal processing for recognition of human frustration. In: *Proceedings of ICASSP98*.
- Francis, W. N. and Kucera, H., 1979. *Brown Corpus Manual* [online]. Department of Linguistics, Brown University. Available from: <http://www.hit.uib.no/icame/brown/bcm.html> [Accessed 27 Jan 2003].
- Franklin, S. and Graesser, A., 1996. Is it an agent or just a program? A Taxonomy for Autonomous Agents. In: *Proceedings of third international workshop on agent theories architectures and languages*. New York: Springer.

- Frijda, N., 1986. *The Emotions*. New York: Cambridge University Press.
- Gaughan, A. P., 1986. *An Analysis of the Impact Of Published News On Stock Prices And Stock Trading Volume*. Thesis (PhD). City University of New York.
- Gordon, T., 1996. *Saussure for beginners*. HarperCollins Publishers.
- Gross, R., McIlveen, R., Coolican, H., Russell, J. and Clamp, A., 2000. *Psychology: a new introduction*. Germany: Hodder and Stoughton General Division.
- Hayes-Roth, B., 2001. Adaptive learning guides. *In: Proceedings of the IASTED, Conference on Computers and Advanced Technology in Education*.
- Hayes-Roth, B., 2003. What makes characters seem life-like? *In: H. Prendinger and M. Ishizuka, eds. Life-like Characters, tools, affective functions and applications*. Germany: Springer.
- He, Q., Hua, Q. and Wang, X. Z., 2002. Learning weights of fuzzy production rules by a MAX-MIN neural network. *Fuzzy systems and mathematics*, 388 - 392.
- Heise, R. D. and Calhan, C., 1995. Emotion norms in interpersonal events. *Social Psychology Quarterly*, 58, 223 - 240.
- Helander, M. G., 1988. *Handbook of Human-Computer Interaction*. Amsterdam: Elsevier science publications.
- Hilitz, S. R. and Turoff, M., 1978. *The network nation: Human communication via computer*. MA: Addison-Wesley.
- Horton, D. and Wohl, R. R., 1956. Mass communication and para-social interaction: Observation on intimacy at a distance. *Psychiatry*, 19, 215 - 229.
- Izard, C.E., 1971. *The face of Emotion*. New York: Meredith.
- Izard, C. E., 1977. *Human Emotions*. New York: Plenum Press.
- James, W., 1878. Remarks on Spencer's definition of mind as correspondence. *The Journal of Speculative Philosophy*, 7 - 22.
- Jensen, C., Farnham, S., Drucker, S. and Kollok, P., 2000. The effect of communication modality on cooperation in online environments. *In: Proceedings of CHI2000*, The Hague, Netherlands.
- Johnson, L. W., Rickel, J. and Lester, J. C., 2000. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International journal of artificial intelligence in education*, 11, 47 - 78.

- Kshirsagar, S. and Magnenat-Thalmann, N., 2002. A multilayer personality model. *In Proceedings of 2nd International Symposium on Smart Graphics*, ACM Press, 107 - 115.
- Kim, H., KoY, P. S. and Seo, J., 1999. Informal requirements analysis: supporting system for human engineer. *In: Proceedings of Conference on IEEE-SMC99*, 3, 1013 -1018.
- Lange, C. G. and James, W., 1922. *The Emotions (Volume I)*. Baltimore: Williams and Wilkins Co.
- Lea, M. and Spears, R., 1991. Computer-mediated communication, de-individuation and group decision-making. *International journal of man-machine studies*, 34, 283-301.
- Leech, G., 1997. *A brief users' guide to the grammatical tagging of the British National Corpus* [online]. Available from:
<http://info.ox.ac.uk/bnc/what/gramtag.html> [Accessed 27 Jan 2003].
- Liu, H., Lieberman, H, and Selker, T., 2003. A model of textual affect sensing using real-world knowledge. *In: Proceedings of the 7th International Conference on Intelligent User Interfaces*, 125-132.
- Liu, K.C., 2000. *Semiotics in Information Systems Engineering*. Cambridge University Press, Cambridge.
- Lotfi, Z., 1965. Fuzzy sets. *Information and Control*, 8, 338 - 353.
- Low, A. W. and Repin, V. D., 2002. The psychophysiology of real-time financial risk processing. *Journal of Cognitive Neuroscience*, 14, 323-339.
- Magnenat-Thalmann, N., Primeau, E. and Thalmann, D., 1998. Abstract muscle action procedures for human face animation. *The visual computer*.3 (5), 290 - 297.
- McCracken, D. D. and Wolfe, R. J., 2004. *User-centred website development: a human-computer interaction approach*. UK: Pearson publishing.
- Mizumoto, M. and Zimmerman, H. Z., 1982. Comparison of fuzzy reasoning methods. *Fuzzy Sets and Systems*, 8, 253 - 283.
- Morris, J. S., Ohman, A. and Dolan, R. J., 1998. Conscious and unconscious motional learning in the human amygdale. *Nature*. 393 (6684), 467 - 470.
- Mountford, S. J. and Gaver, W. W., 1990. Talking and listening to computers, *In: B. Laurel, ed. The art of human-computer interface design*, Addison-Wesley, 319 - 334.
- Murtadha, M., 1975. The problem of contradiction in islamic philosophy. *al Tawhid Islamic Journal*.

- Nakatsu, R., Nicholson, J., and Tosa, N. 1999. Emotion recognition and its application to computer agents with spontaneous interactive capabilities. *In: ICMCS*, 2, 804 - 808.
- Nass, C., Steuer, J., Henriksen, L. and Dryer, D. C., 1994. Machines, social attributions, and ethopoeia: Performance assessments of computers subsequent to "self-" or "other-" evaluations. *International Journal of Man-Machine Studies*.
- Nass, C. and Sundar, S., 1994. Is human-computer interaction social or parasocial? *Paper Submitted to Human Communication Research*.
- Nelson, D., 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59, 349 - 370.
- Nielsen, J., 2000. *Designing web usability*. Indianapolis: New Riders.
- Nielsen, J. and Tahir, M., 2002. *Homepage usability: 50 websites deconstructed*. Indianapolis: New Riders.
- Nye, J. L. and Brower, A. M., 1996. *What's social about social cognition?* London: Sage Publications.
- Ortony, A., 2001. On making believable emotional agents believable. *In: R. P. Trapple, ed. Emotions in humans and artefacts*. Cambridge: MIT Press.
- Ortony, A., Clore, G. and Collins, A., 1988. *The cognitive structure of emotions*. UK: Cambridge University Press.
- Paiva, A., 2000. *Affective interactions, towards a new generation of computer interfaces*, New York: Springer.
- Parke, F. I. and Waters, K., 1996. *Computer Facial Animation*. MA: Wellesley.
- Picard, R. W., 1997. *Affective computing*. MA: The MIT Press.
- Preece, J., Rogers, Y. and Sharp, H., 2002. *Interaction Design: Beyond human-computer interaction*. UK: John Wiley and Sons.
- Prendinger, H., Saeyor, S. and Ishizuka, M., 2003. MPML and SCREAM: Scripting the bodies and minds of life-like characters. *In: H. Prendinger and M. Ishizuka, eds. Life-like Characters, Tools, Affective Functions and Applications*.
- Rice, R. and Love, G., 1987. Electronic emotion: socioemotional content in a computer-mediated communication network. *Communication Research*, 14 (1), 85-108.
- Riding, R. J., 1991. *Cognitive styles analysis users' manual*. Birmingham, Learning and Training Technology.

- Riding, R. J., 1998. *Cognitive styles analysis*, Learning and Training Technology. Birmingham, UK.
- Riding, R. and Cheema, I., 1991. Cognitive styles – an overview and integration. *Educational psychology*. 11, 193 - 215.
- Riding, R. J. and Sadler-Smith, E., 1992. Type of instructional material, cognitive style and learning performance. *Educational Studies*, 18, 323 - 340.
- Riesenhuber, M. and Poggio, T., 2000. Computational models of object recognition in cortex: a review. *In: CBCL Paper #190/AI Memo #1695*. Massachusetts Institute of Technology, Cambridge, MA, August.
- Riseberg, J., Klein, J., Fernandez, R. and Picard, R. W., 1998. Frustrating the user on purpose: using biosignals in a pilot study to detect the user's emotional state. *In: CHI late-breaking*, March 1998.
- Robinson, I., 1975. *The new grammarians' funeral*. UK: Cambridge University Press.
- Ronald, S., Liu, K.C., Hafkamp. M. and Ades, Y., 2000, *Understanding the Role of Signs and Norms in Organisations, - a semiotic approach to information systems design*, *Journal of Behaviour and Information Technology*, 19(1), 15-27.
- Rosis, De. F., Carolis, B. and Pizzulito, S., 1999. Software documentation with animated agents, *In: Proceedings 5th ERCIM workshop on user interfaces for All*.
- Rousseau, D. and Hayes-Roth, B., 1998. A social-psychological model for synthetic actors, *In: Proceedings of 2nd International Conference on Autonomous Agents*. New York: ACM Press, 165-172.
- Rubin, A. M. and Perse, E. M., 1987. Audience activity and television news gratifications. *Communication Research*, 14, 58-84.
- Russell, S. and Novig, P., 2002. *Artificial Intelligence: A Modern Approach*. MA: Prentice Hall.
- Rutterford, J., 1993. *Introduction to Stock Exchange Investment*. 2ND Ed. UK: Macmillan Press.
- Schachter, S. and Singer, J. E., 1962. Cognitive, social and physiological determinants of emotional states. *Psychological Review*, 69, 379-399.
- Schachter, S., 1971. *Emotion, obesity, and crime*. New York: Academic Press.
- Scherer, K. R., 1988. Criteria for emotion-antecedent appraisal: A review. *In: V. Hamilton, G.H. Bower, and N.H. Frijda, Eds. Cognitive perspectives on emotion and motivation*. Dordrecht: Nijhoff, 89–126.

- Schuyler, W. H., 2003. *Reading statistics and research (with research navigator)*. 4nd ed. London: Allyn and Bacon.
- Scott, B. and Nass, C., 2002. Emotion in Human-Computer Interaction. *In: The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*. 81-96.
- Shneiderman, B., 1998. *Designing the user interface: strategies for effective human-computer-interaction*. 3rd. MA: Addison-Wesley.
- Short, J., Williams, E. and Christie, B., 1976. *The social psychology of telecommunications*. London: John Wiley and Sons.
- Simon, H., 1967. Motivational and emotional controls of cognition. *Psychological Review*, 74 (1), 29-39.
- Staller, A., Sloman, S. A. and Ben-Zeev, T., 2000. Perspective effects in nondeontic versions of the Wason selection task. *Memory and Cognition*, 28 (3), 396-405.
- Steinfeld, C., 1985. Dimensions of use of electronic mail in organizations, *In: J. Pearce and R. Robinson. eds. Proceedings of the Academy of Management*.
- Tennant, M., 1988. *Psychology and Adult Learning*. UK: Routledge.
- Tognazzini, B., 1990. Consistency. *In: B. Laurel, ed. The art of human-computer interface design*. CA: Addison-Wesley. 75 - 77.
- Tompkins, S. S. and Messick, S., 1963. *Computer simulation of personality*. New York: Wiley.
- Trappl, R. and Petta, P., 1997. *Creating personalities for synthetic actors*. New York: Springer.
- Trappl, R. Petta, P. and Payr, S., 2003. *Emotions in humans and artefacts*. MA: The MIT press.
- Turkle, S., 1984. *The second self: Computers and the human spirit*. New York: Simon and Schuster.
- Vacca, R. J., 1995. The outer limits - virtual reality on the Internet. *Internet World*, 6 (3), March.
- Velasquez, J., 1997. Modeling emotions and other motivations in synthetic agents. *In: Proceedings of the AAAI Conference 1997*. Providence, RI. 10-15.
- Vyzas, E. and Picard, R. W., 1999. Offline and online recognition of emotion expression from physiological data. *In: Workshop on emotion-based agent architectures, third international conference on autonomous agents*, May 1, 1999. Seattle, WA. 135-142

- Wagner, A., 1990. A Day in the life of an interface designer. *In: Laurel. B, ed. The art of human-computer interface design*, CA: Addison-Wesley, 79 - 84.
- Watson, D. and Clark, L. A., 1984. Negative Affectivity: The disposition to experience aversive emotional states. *In: Psychological bulletin*, 96, 465-490.
- Whaley, E. R., 2002. On the Return and Risk of the CBOE BuyWrite Monthly Index. *Journal of derivatives*, 10, 35-42.
- Windweaver, 1997. *Smileys and emoticons for email and IM* [online]. Available from: <http://www.windweaver.com/emoticon.htm> [Accessed 2 July 2003].
- Winograd, T. and Flores, C., 1987. *Understanding computers and cognition: A new foundation for design*. MA: Addison-Wesley.
- Witmer, B. G. and Singer, M . J., 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence*, 8 (5), 560-565
- Wolberg, G., 1990. *Digital Image Warping*. CA: IEEE Computer Society Press.
- Woods, S. and Arps, J, 1996. Float analysis. *Technical analysis of stocks and commodities*, 12.
- Xiong, R., 2002. *Visualizing information spaces to enhance social interaction*. Thesis (Ph.D). Department of Electrical Engineering and Computer Science, MIT.
- Zaichkowsky, J. K., 1986. The Emotional side of product involvement, *In: P. Anderson and M. Walendorf, eds. advances in consumer research*. Association for Consumer Research, Provo UT, (14), 32-35.
- Zhang, Y., 2001. *Image Morphing* [online]. Oxford University. Available from: www.fmrib.ox.ac.uk/~yongyue/morphing.html, [Accessed 27 Jan 2003].