EMOTIONAL RECOGNITION IN COMPUTING

A thesis submitted for the degree of Doctor of Philosophy

by
Lesley Ann Axelrod

School of Information Systems, Computing and Mathematics
Brunel University
September 2009

Abstract

Emotions are fundamental to human lives and decision-making. Understanding and expression of emotional feeling between people forms an intricate web. This complex interactional phenomena, is a hot topic for research, as new techniques such as brain imaging give us insights about how emotions are tied to human functions. Communication of emotions is mixed with communication of other types of information (such as factual details) and emotions can be consciously or unconsciously displayed. Affective computer systems, using sensors for emotion recognition and able to make emotive responses are under development. The increased potential for emotional interaction with products and services, in many domains, is generating much interest. Emotionally enhanced systems have potential to improve human computer interaction and so to improve how systems are used and what they can deliver. They may also have adverse implications such as creating systems capable of emotional manipulation of users. Affective systems are in their infancy and lack human complexity and capability. This makes it difficult to assess whether human interaction with such systems will actually prove beneficial or desirable to users. By using experimental design, a Wizard of Oz methodology and a game that appeared to respond to the user's emotional signals with human-like capability, I tested user experience and reactions to a system that appeared affective. To assess users' behaviour, I developed a novel affective behaviour coding system called 'affectemes'. I found significant gains in user satisfaction and performance when using an affective system. Those believing the system responded to emotional signals blinked more frequently. If the machine failed to respond to their emotional signals, they increased their efforts to convey emotion, which might be an attempt to 'repair' the interaction. This work highlights how very complex and difficult it is to design and evaluate affective systems. I identify many issues for future work, including the unconscious nature of emotions and how they are recognised and displayed with affective systems; issues about the power of emotionally interactive systems and their evaluation; and critical ethical issues. These are important considerations for future design of systems that use emotion recognition in computing.

Acknowledgements

Exploring emotion in computing has been an amazing, roller coaster ride, sometimes exhausting and frustrating, often wonderfully intriguing and exhilarating and always a life changing, emotional, learning experience. Thanks to Tony for support and for all the music in my life; to good friends like Fran and Sue for saying ‘do it’; to the dog for carrying on taking me for walks (just as if I wasn’t doing it); and to my children, grandchildren and wider family and friends for their patience when I have been a present absence or an absent presence due to the demands of thesis work.

I owe much to my training in human communication science and experience working as a speech and language therapist. In particular, tutors at NHCSS (now UCL), and inspirational therapists like Rosemary Clarke and Jenny Gower, Judy Chaloner and Lindsay Doidge, and not forgetting Ruth Coles, Director of Action for Dysphasic Adults, (now Speakability) who widened the scope of my interests and ambition. I must thank UCL twice over, for training me first as a therapist and later in HCI. Special thanks to John Long, Rachel Benedyk, staff and students from the millennium HCI-E class at UCL who made me welcome to the world of HCI and altered my path in life. Thanks to academic mentors, especially Karen Bryan, at UCL and Surrey, and Geraldine Fitzpatrick at Sussex whose unfailing support gave me confidence.

Thanks to EPSRC for the project grant (R81374/01) that enabled the work to take place, and for opportunities to attend conferences where I was immersed in such interesting work from around the world, and got the chance to meet inspiring researchers and /or to hear about their work, including Cornell, CEmCom group and Ros Picard from MIT.
I was fortunate to undertake this project against a backdrop of much dynamic and exciting research involving social aspects of computing and affective computing that inspired and challenged me, so I thank all those who contributed to that. Particular thanks go to the many contributors to the European Humaine Network, which provided much inspiration, including Nadia Berthouze from UCL and Kristina Höök at SICS.

Thanks to the staff from Brunel who have added to my experience and support: Ray Paul who created a vibrant department; Steve Love who co-supervised; DISC management including Guy Fitzgerald and Rob Macredie; admin team including Julie Whitaker, Ela Heaney and Shirley Hatch; and the technical support team, especially John Park and Jeremy Baxter. Thanks most of all to Kate Hone for empowering me to carry out the work and inspiring and supporting and supervising throughout the process.

Declaration / Date of Submission

I hereby declare that this thesis is my own work and effort and that it has not been submitted anywhere for any award. Where other sources of information have been used, they have been acknowledged.

Lesley Ann Axelrod
30th Sept 2009

Content

Page

1 Introduction to Emotion Recognition in Computing 1
1.1 Introduction 1
1.2 Definition of Emotion 2
1.3 Emotion and the Individual 4
1.3.1 Communication of Emotion 5
1.4 The Wider Emotional Climate 7
1.4.1 The Rise of Emotional Products 9
1.4.2 Emerging Applications 10
1.5 Challenges of Emotion recognition in Computing 12
1.6 Scope of Work, Objectives and Aims 14
1.7 Contribution 16
1.8 Plan of Thesis 17
1.9 Summary 18

2. Affective Literature Review 19
2.1 Introduction 19
2.2 Background Emotion Research 19
2.2.1 Theoretical Perspectives 19
2.2.2 Studies of the Brain 21
2.2.3 Models of Emotion 25
2.3 Human Communication and Emotion 34
2.3.1 Multimodal Human Communication 34
2.3.2 Variables 40
2.4 Terminology 49
2.4.1 Difficulties of definition 49
2.4.2 Definitions Associated with Affective Computing 52
2.5 Affective Computing Applications 55
2..5.1 Automatic Emotional Detection Systems 55
2.5.2 Types of Applications 58
2.5.3 Limits of Computer Recognition and Display of Emotion 63
2.6 Human-Computer Interaction and Emotions 65
2.6.1 Attitudes to New Media 65
2.6.2 Ethical Considerations 66
2.7 Summary 69
2.7.1 Summary of Implications and Issues 69
2.7.2 Overall Aim and Specific Objectives 72

3. Methodology 73
3.1 Introduction 73
3.2 Experimental Design 73
3.2.1 Rationale 73
3.2.2 Method 73
3.2.3 Disadvantages 74
3.2.4 Advantages 75
3.2.5 Conclusions on Experimental Approach 76

3.3 Wizard of Oz Method 79
3.3.1 Method 79
3.3.2 Advantages 79
3.3.3 Disadvantages 80
3.3.4 Conclusion on Wizard of Oz Method 80
3.4 Observational Methods 81
3.4.1 Methods 81
3.4.2 Disadvantages 82
3.4.3 Advantages 83
3.4.4 Conclusions on Observational Methods 83
3.5 Subjective Measurement Methods 84
3.5.1 Methods 84
3.5.2 Quantitative Methods 84
3.5.3 Qualitative Methods 86
3.5.4 Combined Methods 90
3.5.5 Conclusions on Subjective Measurement Methods 90
3.6 Tools 91
3.6.1 Affective Application 91
3.6.2 Range of Coding and Analysis Tools 92
3.6.3 Conclusions on Tools 95
3.7 Summary 95

4. Affective Experiment Design 98
4.1 Introduction 98
4.2 Study Design 100
4.2.1 Hypotheses 100
4.2.2 Factorial Design - Independent Variables 102
4.2.3 Dependent Variables 103
4.2.4 Participants 104
4.2.5 The Game 104
4.2.6 Equipment 108
4.2.7 WOZ Procedure 109
4.2.8 General Analysis 112
4.2.9 Coding Procedure 112
4.2.10 Reliability of the Method 117
4.3 Summary 121

5**. Affective Experiment Results** 123
5.1 Introduction 123
5.2 Initial Results from Coding Procedure 123
5.3 Overall Results 127
5.3.1 Task performance 127
5.3.2 Subjective Satisfaction 128
5.3.3 Valence 129
5.3.4 Arousal 130
5.3.5 Behaviours 131
5.3.6 Qualitative Data 135
5.4 Discussion 140
5.5 Summary 143

6. Follow-on Experiment 146
6.1 Introduction 146
6.2 Experiment Rationale and Design 146
6.2.1 Research Questions 147
6.2.2 Preparatory Work 148
6.2.3 Pilot Behaviour training 149
6.2.4 Study Design and Procedure 150
6.3 Results 152
6.4 Discussion 153
6.5 Summary 159

7. Discussion 160
7.1 Introduction 160
7.2. Complexity of Emotion Related Behaviours 160
7.3 Intentional Affect or 'Plain' Communication? 166
7.4 Training Emotional Behaviours 167
7.5 Feelings Confounding Feelings 168
7.6 Study Design Issues 168
7.6.1 Preparatory Work 168
7.6.2 Subjective Requirements 169
7.6.3 Using a WOZ Scenario 170
7.6.4 Baseline Behaviours 171
7.6.5 Normality of Participants 171
7.6.6 Technical Issues 172
7.6.7 Masking 172
7.7 Coding Method and Validation 172
7.8 Design Implications for Affective Systems 176
7.8.1 Terminology 176
7.8.2 Ethical Issues 177
7.8.3 Tools and Methods 178
7.9 Summary 178

8. Conclusions and Contributions 180
8.1 Introduction 180
8.2 Contributions 180
8.2.1 Objectives Met 181
8.2.2 'Affectemes' Coding Scheme 183
8.2.3 Wider Issues 184
8.2.4 Issues for Future Work 187
8.2.5 Contributions Chart 188
8.3 Future Work 189
8.4 Summary 194

9. Appendices 197

A. List of publications and dissemination arising from this work 198

B. EPSRC report on project 202

C. Draft final version of CHI'05 conference paper: 203

Axelrod, L.; Hone, K. 2005 E-motional Advantage: Performance and Satisfaction Gains with Affective Computing. *Proceedings of CHI 2005* 'Technology, Community, Safety' Conference on Human Factors in Computer Systems, Portland, USA.

D. Draft final version of Behaviour and Information Technology journal paper: 210

Axelrod, L.; Hone, K. 2006 Affectemes and allafects: a novel approach to coding emotional expression during interactive experiences. Behaviour & Information Technology, special issue on Empirical studies of the User experience Vol. 25, No. 2, 159 – 173.

E. Advert, Game Details, Participant Instructions, and Consents, Pre and Post Experiment Scales and Question Sheets. 232

E. Quantitative Analysis 250

F. Qualitative Analysis 282

10. References 295

Tables

		Page
Table 1.	MindReader categories of emotion from 'Mind Reading Emotions Library User Guide version 1.2' P16.	3
Table 2.	Ortony and Turner, (1990) 'What's basic about basic emotions?'	27
Table 3.	Genera of emotional states (after Scherer 2004) from Humaine deliverable document D3i Emotional life: Terminological and conceptual clarifications	53
Table 4.	2x2 factorial design	103
Table 5.	Second-level transcripts and keywords used.	116
Table 6.	Sample 1 of coding of data	118
Table 7.	Sample 2 of coding of data	118
Table 8.	2 x 2 factorial design and number of rungs completed.	128
Table 9.	2 x 2 factorial design and subjective satisfaction	128
Table 10.	2 x 2 factorial design and valence	129
Table 11.	2 x 2 factorial design and arousal	131
Table 12.	2 x 2 factorial design and number of affectemes	132
Table 13.	2 x 2 factorial design and comparison of no. affectemes in 2 samples	132
Table 14.	2 x 2 factorial design and extent participants believed they showed emotion	133
Table 15.	2 x 2 factorial design and blink rate	134
Table 16.	2 x 2 factorial design and clues given	135
Table 17.	Views on development of emotion recognition systems.	139
Table 18.	Results Summary	152

Figures

		Page
Figure 1.	Illustration of Shaver et al's (1987) Hierarchical Structure of Emotions.	2
Figure 2.	Russell's circumplex.	26
Figure 3.	Plutchik's cone (1980).	26
Figure 4.	The Cogaff model (from Sloman's Cognition and Affect Project website).	29
Figure 5.	Minsky's six structural levels of emotion (from Minsky, (2006) 'The Emotion Machine').	29
Figure 6.	OCC Model (from Ortony Clore & Collins, (1988), The Cognitive Structure of Emotions pp19).	30
Figure 7.	Scherer's component process model from Humaine website	31
Figure 8.	Don Norman's levels of emotional processing (from 'Emotional Design' 2004).	32
Figure 9.	Humunculus	36
Figure 10.	Sample structure of word-ladder game as presented for solution.	106
Figure 11.	Sample structure of word-ladder game with clues and solutions.	106
Figure 12.	Screen shots of Word-ladder application.	106
Figure 13.	WOZ scenario with researcher view on left and participant's view on the right.	110
Figure 14.	Array of emotional expressions of one participant	124
Figure 15.	Variations on a chin dump. Allaffects.	126
Figure. 16.	Interaction plot for ratings of positive valence	130
Figure 17.	Interaction plot for self reported emotion display	134
Figure 18.	Contributions Chart	188

Emotional remarks that inspired….

Once upon a time, computers processed data. ….. Then along came broadband networks, multimedia, and mobile devices, and with them came fun, persuasion, outrage, delight, faith, campaigns, satire, lifelong learning, identity, communities and passion. Now, increasingly, we interact to be, not just to do. Interaction no longer just changes things, it changes people. …….Communication informs, engages and persuades, and thus creates new horizons for all human activity. We have already set sail towards these new horizons. The anchor's up, but we don't quite know where we're bound. ……Bring distant shores into view. Chart new routes to new destinations. Each new route will bring fresh opportunities. Little will remain unchanged - publishing, learning, marketing and politics will all change in the face of disruptive technologies. These changes must be guided by HCI's human perspective and balance. As a result, HCI must be renewed with the fullest understanding of what it is to be human and what our humanity implies for interactive communication in the digital world.

….Issues of emotion, affective response, and inclusive human concerns are exceedingly important in the HCI community. As people become more sensitive to dimensions of products that go beyond traditional aspects of usability, the need to understand and create emotional and aesthetic resonance between people and technology products increases. However, we have yet to discover a shared understanding and develop a shared language for emotion within the context of design. …..The various disciplines involved in Human-Computer Interaction each bring their own theories and languages about emotion to the design and development process.

Gilbert Cockton and Panu Korhonen
Co-Chairs CHI 2003 welcome speech (CHI 2003 website).

"What I'm really saying is that the technologies that are emerging are sufficiently powerful that they can be used to redesign ourselves in the world. And I don't think there's much dispute about that, that using biotechnology we can reengineer our species if we so choose. Using nanotechnology we can do all sorts of amazing things if we so choose and that we will eventually be able to make intelligent machines. I think some people may argue about whether it's thirty years or a hundred years but those technologies are sufficiently powerful that we can reinvent the world and then reinvent ourselves. And all I'm saying is that rather than letting whatever happens happen, we ought to think about what kind of world we want to have. If we have the power to invent it, we ought to take some time and have a discussion about what kind of world we want to have. And the first part of that is to take and have a discussion about how much risk we want to take because these technologies are very risky. They're proponents and no one really denies that anymore."

Bill Joy, Chief Scientist, Sun Microsystems
interview with Peter Robinson, 2001 (Hoover Institute interview).

"It does not do to leave a live dragon out of your calculations, if you live near him."

From J R Tolkien, in The Hobbit,
quoted as a warning in The Age of Spiritual Machines (Kurzweil, 1999).

"Emotions are part of a management system to co-ordinate each individual's multiple plans and goals under constraints of time and other resources. Emotions are part of the biological solution to the problem of how to plan and to carry out action aimed at satisfying multiple goals in environments which are not perfectly predictable. Emotions are based on non-propositional communications which we will call 'emotion signals'. They function both to set the whole system suddenly into a particular mode, and to maintain it tonically in that mode. Emotion signals provide a specific communication system which can invoke the actions of some processors and switch others off."

Towards a Cognitive Theory of Emotions
Oatley and Johnson-Laird (1987).

"Emotion….is a biologically given sense, and our most important one. Like other senses – hearing, touch, smell – it is a means by which we know about out relation to the world and it is therefore critical for the survival of human beings in group life."

The Managed Heart
Arlie Hochschild (1983) p.219.

"The streets of London have their map; but our passions are uncharted. What are you going to meet if you turn this corner?"

Jacob's Room
Virginia Woolf (1922).

"Let's not forget that small emotions are the great captains of our lives, and that these we obey without knowing it."

Letter to Theo van Gogh
Vincent Van Gogh (1889).

"There are moments in life, when the heart is so full of emotion
That if by chance it be shaken, or into its depths like a pebble
Drops some careless word, it overflows, and its secret,
Spilt on the ground like water, can never be gathered together"

The courtship of Miles Standish. VI Priscilla,
Henry Wadsworth Longfellow (1858).

1 Introduction to Emotion Recognition in Computing

1.1 Introduction

This thesis explores the use of emotion recognition in computing, from the user perspective. As computers become ever more ubiquitous and multi-purpose, systems are being developed that attempt to recognise aspects of emotion related behaviours and to respond to these, for example systems designed to improve the user experience or to change user behaviour.

"Emotion" is a large topic area with varied use of terminology and many concepts, theories and models associated with it. The term "affective computing" is now commonly used in the human-computer interaction (HCI) field to embrace systems that address human emotion in some way and this has become a major topic of interest in the community over the last fifteen years. Initially development of affective systems was biased to technologically driven systems with only limited conceptualisation and consideration of human, emotional aspects. User views were (and to an extent still are) difficult to study because systems simply are not good enough to test in realistic situations. The work described in this thesis shows how using Wizard of Oz techniques enables simulation of an affective system and exploration of how users perform with it. The simulation provides a setting that enables collection of very rich qualitative data about the user experience and how users feel about using an affective system. This work shows how very complex and challenging working in the arena of affective computing is, and contributes to the swing towards a more user centred and interactional view of emotion in computing.

This introductory chapter first gives a broad outline of definitions relating to emotion as used by others and explains use of terminology. It then reviews the importance of emotional life to the individual and relates that to the current cultural climate. In order to set the context of this thesis, it goes on to give an idea of the kinds of technologies being developed and it outlines the challenges of emotion recognition in affective systems and of coding affective behaviours. Finally, it describes the subsequent scope of work and the structure of this thesis.

1.2 Definition of Emotion

Emotion is difficult to define, and there is no universally accepted single definition. The word ‘emotion’ is used to refer to both the inner feelings of a person and to their outer displays of those feelings.

Researchers have developed various taxonomies of emotion related words for use in different domains or contexts. Some of these categorise words according to linguistic analysis, (e.g. Oately & Johnson-Laird, 1987), some by scaling of similarity judgements (e.g. Shaver, Schwartz, Kirson & O’Connor, 1987) (see figure 1.), or some by analysing emotion events in various scenarios (e.g. Izard, 1977; Reisenzein, 1995).

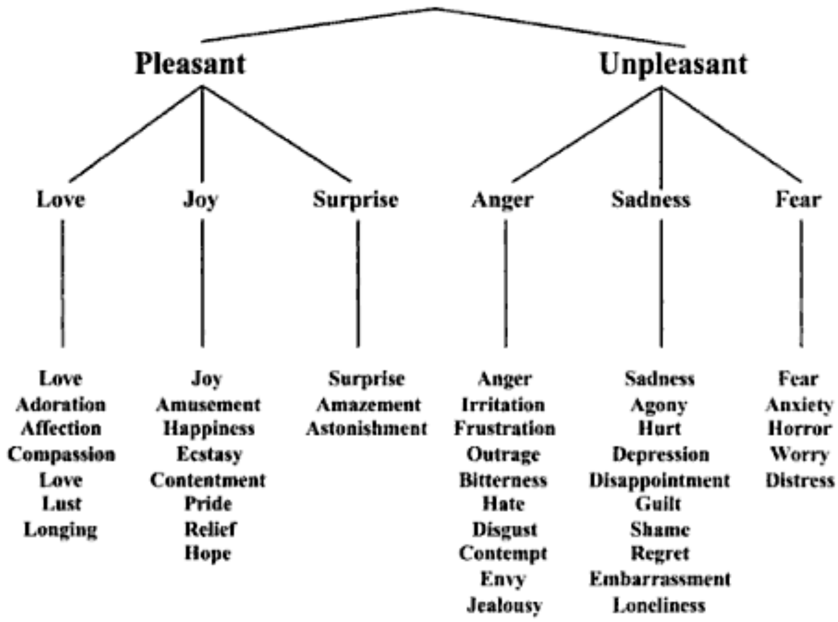


Figure 1. Illustration of Shaver et al’s (1987) Hierarchical Structure of Emotions

Lexicons can be very useful tools, for example, Baron-Cohen used lexicographic analysis of 1150 words (412 concept words and 738 synonyms) from the Microsoft Thesaurus and grouped them into 24 mutually exclusive groups, that formed the basis for his Mind Reading application (Baron-Cohen, 2003; Baron-Cohen, Golan, Wheelwright & Hill, 2004). Mind Reading is an interactive computer-based guide to reading emotions from the face and voice that uses actors to demonstrate 24 complex emotional states (see Table 1.).

Afraid	Angry	Bored	Bothered	Disbelieving	Disgusted
Excited	Fond	Happy	Hurt	Interested	Kind
Liked	Romantic	Sad	Sneaky	Sorry	Sure
Surprised	Thinking	Touched	Unfriendly	Unsure	Wanting

Table 1. MindReader categories of emotion from 'Mind Reading Emotions Library User Guide version 1.2' P16.

Emotion words have many subtle shades of meaning in different contexts (that do not always readily lend to straightforward translation between languages). Baron-Cohen likens his categories to the colour spectrum:

> "We can think of these 24 groups as being similar to the colour spectrum, which is comprised of groups of colours (the Red Group, the Blue Group, and so on), which have shades of colours within them (e.g. shades of blue such as turquoise and aqua-marine are in the Blue Group). Thus, every emotion falls into a specific group (e.g. grumpy and furious are in the Angry Group). When we use a word like "blue" it could refer to a whole group of colours, or it could refer to the clearest, most typical example of a shade of blue, within that group. So it is with emotions. When we use the word "angry" it could refer to a group of emotions, or it could refer to the clearest, most typical example of a shade of anger, within that emotion group." (Baron-Cohen, 2003, P7.)

With emotion labelling, as with colour labelling there is a further complication in that we can never know exactly how another person's perception compares to our own. So for example a colour one person perceives and labels as 'red', might be perceived as 'pink', but labelled as 'red' by another. Or with emotion words, one person's 'miffed' might be another person's 'angry' or 'furious' (Brown & Silzer, 2001). These differences in word use might be due to underlying differences in physical and neurological sensory capacity or they may be due to cultural or idiosyncratic differences in use of words for labelling.

One aim of the Europe-wide Humaine initiative (see Humaine website, 2009) was to establish a vision of a common emotion language. This grew from a preliminary list of 55 emotion terms to a 50 page document that described and defined multifarious aspects of emotion such as Scherer's framework (to differentiate emotions, moods, stances, attitudes and dispositions). In the final consensus, the field was described as:

> "a domain that seems important intuitively, but which is difficult to describe explicitly" and that can be summarised as "felt appraisals, states, connections, expressions and tendencies' and said to be "more or less what Hume meant by 'the heart"(Cowie, 2007 p.50).

Emotion is now generally recognised as a complex human system, involving both body and brain that inter-relates to other internal systems such as cognition and memory, and operates on physiological, neurological and reflective levels (Norman, 2004) (further explored in Chapter Two of this thesis). Despite the difficulties about definition, it is possible and important to consider issues associated with the concept of emotion in computing. Picard uses a striking metaphor comparing emotion with Mount Everest – everyone knows what is meant by the term 'Everest' although it is impossible to say exactly which individual rocks and slopes can be defined as part of the mountain (Picard, 1997). However Picard suggests that we should not let this difficulty with definition prevent us from exploring the rocky terrain.

Within the field relating to 'emotion', I need to establish the boundaries I will observe and the scope for this study. I choose to follow Picard (1997) who uses 'affect' to include the wider domain of emotional terms and I clarify throughout whether I am referring to inner feelings or outward, visible expressions of emotion. The focus of this study is mainly on observable, emotion related behaviours and outward expressions of emotion, as illustrated in the analysis phase of this work.

1.3 Emotion and the Individual

Despite the recognition by philosophers of the importance of emotions to the human condition, historically emotions were often considered a hindrance to rational thought. For example, Plato saw the human soul as consisting of three basic energies, reason, emotion and appetite, that must be ruled by reason for a balanced life. He used an allegory of a chariot (representing the soul on its journey) being pulled in two directions by two winged horses, one black and one white, representing positive and negative passions. The charioteer must use reason and control to steer the chariot (Plato, Jowett's 1999 translation). More recently, our ambiguous views on emotion are revealed by the depiction of fictional characters that are frequently portrayed as superior problem solvers because they *lack* emotion, for example, Dr Spock in Startrek (Gibberman, 1991). Only in recent decades have researchers begun to explore the real value of emotions to mankind and to scientifically study the role emotions play in human development and every day life and well being, for example in forming relationships, pro-social behaviour, or in higher order thinking such as cognition, reasoning or problem solving, motivation, consciousness, memory, learning and creativity (Salovey & Mayer 1990). We use emotions to help us to learn and regulate our behaviour and emotions are now recognised to play an important part in everyday life, including cognition and the laying down and retrieval of memories (Scherer, Wrankin, Tran & Scherer, 2004; Reisberg and Heuer, 2004).

1.3.1 Communication of Emotion

Emotional displays are a normal part of human communication and their use and recognition is fundamental to successful interaction. Like other aspects of human language we seem to have an inbuilt ability to acquire and use it and we seem 'hard wired' for emotional interaction with empathic mirroring behaviours evident even in newborns' behaviour with their attention to and imitation of facial expression (Nadel & Butterworth, 1999). Emotional feelings may be diminished or absent in those who for physical or neurological reasons are unable to imitate. Research with people with known localised brain damage or deficits in brain tissue and function due to accidents or injury have informed research, for example Adolph's study of individuals with damaged amygdala regions of the brain found they had impaired ability to recall emotional stories (compared to stronger recall of emotional compared to non–emotional stories in normal individuals) (Adolph, Cahill, Schul & Babinsky, 1997).

Some basic expressions are said to be universal (Ekman, 1982), and some are inborn, as even blind babies, who cannot see facial expression can still have a 'smile' reflex. Emotions vary with individual or cultural differences (Scollon, Diener, Oishi & Biswas-Diener, 2004). The degree of expertise and exact modes of use and understanding of emotional expressions may vary from one culture or individual to another. In the same way that written language has followed on from spoken, it is arguable that non-verbal means of communication are more original and fundamental in evolutionary terms, than spoken language.

Researchers, such as Bradley and Cahill have found that arousal levels enhance memory (Bradley, Greenwald & Petry, 1992; Cahill & McGaugh, 1995) and Isen found benefits from positive affect (Isen, 2000). Damaiso found that having too little emotion impairs decision making and developed a Somatic-Marker Hypothesis (SMH) which proposes that emotional processes guide behaviour and in particular decision-making (Damasio, 1994). Despite recent work e.g. using functional magnetic resonance imaging (fMRI) scans to try to identify the parts of the brain active during experience of emotion (e.g. Dolcos, LaBarr & Cabeza, 2005), we can never be sure of the exact experiences of inner emotions that individuals perceive and feel. The quality of inner emotional feelings may or may not be effectively communicated between people.

We can display emotions in many different ways (Argyle 1996) and often use more than one means at a time. As with all human communication there is much redundancy built in to systems of emotional expression. Expression can be intentional or unconscious. Sometimes the expression of emotion is inhibited e.g. in an attempt to hide inner feelings. In order to recognise human emotions it is necessary to perceive the signals – to see the facial expression, hear the tone of voice etc. and then to interpret their meaning according to the context and known patterns and rules. Individual skill in both display and recognition of emotions varies (as illustrated in our later section on the inter-rater reliability of coders of emotional behaviours described in chapter 4 and discussed in chapter 7). And of course emotion is not the only thing that humans display – they may use similar displays to communicate all sorts of information or to comply with social rules, without experiencing any underlying emotional experience. For example in Western cultures it is polite to smile when greeting someone, (Kendon & Ferber, 1973) but that does not necessarily reflect an underlying happy state. A downturned lip and the slant of an eyebrow might reflect a state of inner grief, might be a pretence of grief, or might be used in conjunction with joking words to show irony, or might just be due to the physiology of a particular individual. As with all human communication, non-verbal signs leave much scope for error and misunderstanding. Both inner feelings and ways of expressing emotion during human-human interaction may vary, according to variables such as age, gender, cultural norms, mood, temperament, personality and with variation in the intention of an individual to convey messages about themselves or about their situation.

1.4 The Wider Emotional Climate

The emotional state of society has become a political issue and is becoming ever more critical as we face global challenges such as those posed by climate change, the economy and the ticking time-bomb of our ageing population. It is seen as critical for sustainable growth, and with an ageing population it is essential that those of working age are fit to contribute to the economy and for older people to remain functioning well and independently in society for as long as possible (European Commission, 2005) and emotional wellbeing is seen as key to achieve this. Whilst the USA has always incorporated a right to the 'pursuit of happiness' enshrined in their constitution, the UK has now taken measures to address the emotional state of the nation. Following a global trend, (for European examples see Abdallah, Thompson, Michaelson, Marks & Steuer, 2009) and in order to address the challenges of sustainable development of the economy, in 2005, the UK government announced plans for an official "index of wellbeing" This is intended to supplement measures of Gross Domestic Product, as a means to assess the progress of the nation (Newton, 2007). The need for society's health and emotional wellbeing is addressed in Dame Carol Black's review for the UK government's cross-departmental initiative for health, work and well being (Black, 2008). Leisure and exercise, (seen as crucial for tackling the looming obesity crisis (Lear & Palmer, 2008), are counted as important components for emotional wellbeing.

In the information age, with the UK manufacturing industry in decline, our service economy is becoming a 'relationship economy' with emotional work incorporated into business and seen as a commercial necessity. Emotional management and monitoring in the form of ‘emotional labour’ are recognised as central to our working lives and many working roles emphasize the importance of relationships and emotional interaction skills, for example in hospitals ‘named nurses’ are allocated to care for patients. The emotional labour required in work, was first highlighted by Hochschild (1983), in her studies of the emotional management carried out by airhostesses. More recent work has extended this concept to health, leisure and home settings (Smith, 1992; Rojek, 2009). It is recognised that emotions can give commercial edge to business, by improving relationships and communications, diffusing conflict, improving customer services such as complaints handling and by improving customer retention. From ‘greeters’ paid to welcome us into shops, to vast call centres, an ‘experience culture’ is emerging, where management and manipulation of emotions is key (Bunting, 2004). At the same time computing systems are increasingly embedded in working practices and mediate working transactions, for example for home working and collaborative work across distances with colleagues or clients, such as online shopping, banking, call centres etc. or as tools, for example for health monitoring and recording.

On the home front, emotional well being has become a focus as we struggle to cope with the changing structure of society. In the past people mostly communicated face to face or via hand written letters within a close-knit structure of people, such as lifelong work colleagues, the immediate nuclear family, or friends and neighbours living geographically nearby. Now individuals often struggle to maintain complex extended relationships due to broken and or extended family set ups and with an increasingly dynamic and mobile population. At the same time, new technologies such as mobile phones and social networking afford new means of communicating regardless of geographical boundaries and there has been a rapid uptake and growth in electronic communication for a wide range of social purposes (Ofcom 2007, 2009).

With increasing numbers of dependent elderly it is crucial that younger members of society are fit to support them financially and emotionally. The need to master emotional competence is recognised as essential for successful integration in our society and for sustainable development to face the challenges of the 21st century.

Consequently teaching of emotional skills has attained a high profile and emotional intelligence components have been added to the core personal and social section of the National Curriculum for all UK school children to benefit from teaching of emotional skills:

> "Personal wellbeing helps pupils explore .... and reflect on and clarify their own values and attitudes. They identify and articulate feelings and emotions, learn to manage new or difficult situations positively and form and maintain effective relationships with a wide range of people. Personal wellbeing makes a major contribution to the promotion of personal development." (Qualifications and Curriculum Authority, 2009).

1.4.1 The Rise of Emotional Products

Emotion has long been a topic of interest for philosophers and others interested in the human condition (see Oatley, 2004). We have traditionally used a number of tools to reflect on, help interpret and express emotional feelings, for example artists, musicians and storytellers have always intuitively played upon our feelings and emotion. Emotion has always been a part of business processes, for example with deals relying on relationships and emotional sales pitches.

The current social climate has resulted in a surge of cross disciplinary research into emotions from diverse fields such as social science, anthropology, biology, neurology, psychology, linguistics, human communication science, information systems, Artificial Intelligence (AI), etc. Emerging understandings about the structure and role of emotions is proving useful in many aspects of business and product development from toys to business processes. Norman suggests that making something more pleasant makes it easier to use so that emotional design and 'fun' products are becoming requisite. (Norman, 2004; Blythe, Monk, Overbeeke & Wright, 2003).

At the same time as we place a greater emphasis on our emotions, we have a stream of new technologies or novel arrangements of older ones, with potential uses for emotional manipulation, that are integral to our daily lives and that require HCI and design expertise, as summarised by Gilbert Cockton and Panu Korhonen (CHI 2003 General Conference Co-Chairs):

> "Along came broadband networks, multimedia, and mobile devices, and with them came fun, persuasion, outrage, delight, faith, campaigns, satire, lifelong learning, identity, communities and passion. Now, increasingly, we interact to be, not just to do. Computers are no longer just data crunchers. Individuals and communities used them first for work – but now for education, fun, entertainment and communication in everyday life with ambient, ubiquitous, mobile, immersive and multimodal applications." (Cockton, & Korhonen, 2003).

1.4.2 Emerging Applications

The work described in this thesis evolved in the context of very early emerging applications using affective technologies with limited capabilities. Innovative and imperfect affective technologies were attracting media attention and firing the public imagination. Emotional machines seemed to herald a brave new world – that might be either utopian or dystopian. Professionals in diverse domains such as e-learning, e-commerce and gaming recognised potential for computer systems that consider and involve users' emotions in overall design of products.

Understanding the 'right' degree of emotional design and incorporation of the 'right' level of emotional recognition and response is critical for ethical, effective and commercially successful products. Emotion has long been recognised and used in many of life's arenas. Businesses capitalise on emotional techniques and various media industries that are highly verbally and visually based such as advertising, publishing etc. have long understood certain rules for emotional interaction, methods for manipulating the emotions and ways to use those methods to persuade their target audience. Emotional manipulation is designed into business processes to some degree, for example by using a friendly communication style to promote customer satisfaction; by using attractive product design to attract customers (e.g. see Cockton, 2009 on Kansei and design); by uses of persuasion to achieve sales (e.g. see ChangingMinds.org, 2010); or to change or 'nudge' behaviour (e.g. Thaler & Sunstein, 2009). In the home, products to promote a rich emotional life and balance are considered desirable. In the health arena, persuasive systems are being developed that encourage behaviour change towards a healthier lifestyle (e.g. Consolvo, Klasnja, McDonald, & Landay, 2009).

With affective technologies, recognition of emotional states is the aim in some applications. For example police have been carrying out trials of use of automatic behaviour recognition using surveillance cameras to identify people at risk of offending (University of Southern California, 1997). There are projects to recognise affective states from posture, for example helping doctors to recognise pain in order to enhance medical care (Bianchi-Berthouze, 2008).

In other areas some representation of expression of emotion by a machine is desired, and this might be to regulate, counter or enhance a particular emotional state. For example video games (a large market sector) might be more engaging with emotionally enhanced expressive avatars (e.g. Ortony, 2002) and it is suggested that emotional enhancement might make companion robots more acceptable in social care (e.g. Breazeal, 2000; Bickmore & Cassell, 2001; Koay, Syrdal, Walters & Dautenhahn 2007). There is scope for enhancing applications by using embodied characters, likeable or persuasive avatars or agents, in every arena from sales, to health behaviour change, to gambling games (Creed & Beale, 2006; Rehm & Wissner, 2005).

Systems that both recognise and react to emotions have potential. We know that stress and the way an individual responds to and manages that stress, is related to a number of health issues e.g. in asthma (Rosenkranz & Davidson, 2009) and there are numerous applications that aim to decrease stress and enhance positive affect by recognising and adjusting to the mood of a person. For example smart homes or office environments that adjust music or lighting to individual mood or state of mind.

There has been an emphasis on how technology might help in social situations where recognition and management of emotions is important. We all have varied skills in this department that might vary within an individual from time to time, for example due to fatigue. A system that could recognise and discretely prompt someone as to the state of a communication partner might be useful in many situations. For those who find emotional recognition skills and exchange rules difficult e.g. those on the autistic spectrum, there is particular scope for technologies to aid them. Simply wearing a wristband that gives feedback on arousal level has been found to help individuals to reflect and to decode their own responses to social situations (Poh, Swenson & Picard 2010). It may be possible to develop wearable devices that help identify emotions of others in real time. Using interactive systems to teach emotional skills to those who have deficits is another area of research (Baron-Cohen, Golan, Chapman & Granader, 2007).

Manipulation of mood is another potential area. For example in situations where problem solving is important it might be advantageous for people to be in a happy state as a number of studies show that people solve difficult problems better when they're happy than when they're neutral or angry (Isen, 2000). In the classroom there is some evidence for negative as well as positive affect playing a part in motivating students (Mathews, 1996) and video games are designed to contain emotional arcs that take players on an enjoyable roller coaster ride from one state to another (Zimmerman & Salen, 2003).

There is a darker side to the rise of emotional technologies and the ways they could potentially be used that has so far been little explored but must be considered. For example, Creed and Beale (2008) describe the potential of systems designed to use emotional manipulation for nefarious purposes and the resulting ethical issues.

1.5 Challenges of Emotion Recognition in Computing

Affective systems are a rapidly evolving field, where the needs of users are far from fully established. Affective computing is identified as a challenging area for HCI practitioners, with a need to advance knowledge in this area (Cockton, 2004).
There are significant challenges for affective systems. The technical challenge of building systems that can recognise or show aspects of emotion is an important consideration but development resources such as effort, time and money might be wasted if we do not take a user centred approach. Early development of systems has

been driven by what is technologically possible rather than by user preferences. It is not yet known what people will reject, prefer to use, or find most useful. We must consider issues around user needs, requirements, acceptability, accessibility and effective usability in order to ensure successful design and uptake of systems and subsequent successful interaction and use of them. We need to identify issues and guidelines for affective systems design.

Emotion detection technologies are currently far from the level of human capability in detecting and responding to emotion during interaction. Advances in technology and development of systems is rapid and it has been suggested that affective systems are currently at the stage that interactive voice systems were at some fifteen years ago and enthusiasts suggest that within a decade affective systems may become truly usable. Forecasters predict that affective recognition systems could potentially complement or exceed human capability in time and one sector believes that humans with digitally enhanced capabilities will be the norm in the future (Kurzweil, 1999). However, predictions often fail, and despite early claims, present day speech recognition systems are still limited and uptake has not been as rapid and universal as many predicted. Experiments that explore affective technologies are badly needed for us to ‘get it right’ as we go on to develop useful systems. In this early stage, when we do not yet have fully operative and effective affective systems, we will have to use techniques such as using simulated systems to improve our understanding of human computer interaction with such systems.

Affective computing can learn and benefit from early work in human factors and from work during the rise of automatic speech recognition systems. Speech systems developed rapidly and improved due to technological innovations, and became a valid option for use with end users but had continuing issues around liking and acceptability. Such systems, relying on spoken language, were pioneered for those with disabilities such as visual impairment but now systems are found in many domains in an effort to give companies a commercial edge e.g. in sales and telemarketing. In comparison to speech interfaces, interfaces that recognise or respond to emotional signs may be easier, more intuitive to use and better accepted. It is important to focus on user centred design, to use empirical measures and iterative design (Gould & Lewes, 1985). Methods such as the ‘Wizard of Oz’ that proved useful in early human-computer studies and emphasis on exploring communication and how words and language are used and understood (Chapanis, 1965; Kelley, 1984) might also be useful in this new area.

Telemetry and bio-sensing systems to detect emotion related behaviours have been in use for some time, and much work is being carried out to recognise emotions through measurable physiological data on muscle movements, vocal pitch patterns etc. Work in this field is now beginning to combine multimodal signals eg spoken and visual or information about posture and dialogue (e.g. D’mello & Graesser, 2007). And sensor technologies are advancing so that we can collect physiological data relatively unobtrusively via wearable technologies (Peter, Ebert & Beikirch, 2005) or remotely placed sensors (Prance, Beardsmore-Rust, Watson, Harland, & Prance, 2008). Automatic, real time analysis of physiological data to identify affective states is becoming a reality e.g, Noldus ‘Facereader’ (2009) can give probability of complex emotional states (based on el Kaliouby & Robinson 2004).

There is evidence that users tend to treat new media like real people and real things, (Reeves & Nass, 1996) and this may be even truer for systems that incorporate affective aspects, but the exact mechanism by which this functions, as well as potential benefits, strengths and weaknesses and ethical considerations – all require further exploration.

1.6 Scope of Work, Objectives and Aims

This introductory chapter describes my original research interest in the potential for emotion recognition technology to improve the quality of human-computer interaction and explores the broad field of emotion recognition in computing. Arising from my introductory reading, I developed an interest in establishing what emotions people display during interaction with existing technology; how they display them and what they mean; if there are clear patterns of display with different technologies; and if and how such patterns vary according to variables such as age, gender, background, context etc. and how useful recognising and/or responding to emotional displays might be in the field of human computer interaction.

Chapter Two goes on to describe my review of the literature and the state of the art. The scope, aims and objectives of this study were determined by the literature survey that clearly shows a number of issues to consider about developing computer systems that involve emotion. Existing research is so far limited, leaving wide areas still to explore such as:

- user attitudes to affective computing;
- what affective displays users initiate or use in response to various types of computer applications, including standard graphical user interfaces, let alone empathic or affective displays;
- what users actually prefer in the way of affective response by an application and how this varies in different domains;
- how emotion recognition technology can improve the quality of human-computer interaction;
- implementation issues.

These questions have not been fully addressed because the current state of technology with systems is simply not yet good enough to be implemented in realistic applications. Subsequently, my scope is to focus on the user perspective of affective systems and carry out empirical and qualitative work in order to contribute to knowledge about human-computer interaction issues.

After exploring the broad field of emotion recognition in computing, and following my review of the literature and state of the art (Chapter Two), I developed specific objectives for my study:

1) To establish the extent to which people will naturally express emotions when they know they are interacting with an emotion-detecting computer. As we do not yet have working affective systems, we do not know if users will be willing to show emotions to a computerised system, if they will act normally, or hide their feelings.

2) To identify the conditions under which the application of emotion detection can lead to improvements in subjective and/or objective measures of system usability. It is very important that effort and costs for design of affective systems and products is

directed appropriately. It should be applied in areas where it can make a real contribution to individuals or to society as a whole. It is only by increasing understanding about settings and aspects of affective systems to see where they might add to usability (or other positive aspects such as likeability) that we can achieve this.

3) To establish if and to what extent participants adapt their behaviours during interaction with an apparently affective system. As we do not yet have working affective systems we do not know how or even if people will adapt their behaviours. They might hide their feelings or they might overly display how they are feeling in this novel situation. We won't know what aspects of design are important in development until we know more about user behaviour and we can't study user behaviour until we have developed systems to test.

In Chapter Three I discuss possible methods, identify a need for research, form specific aims and objectives to be achieved from the research and identify expected contributions to the academic society. To address my objectives, I chose empirical experimental design and a Wizard of Oz method in an experimental procedure (described in Chapter Four) to establish user preferences and gains (described in Chapter Five). These results, along with the insights from a follow on experiment (described in chapter Six) helped to identify issues for discussion (Chapter Seven) leading to conclusions and issues for further consideration (Chapter Eight), that are relevant to development of systems that use emotions in computing.

1.7 Contribution

My contributions include illustrating the complexity and many challenges of working within the domain of emotion recognition in computing that must be addressed in order to incorporate emotional considerations into the design of computer based systems. This work adds to the evidence on the user perspective and adds to methodologies for evaluating affect and identifies issues for further investigation. This contribution is important because affective computing has huge potential both for commercial potential and to enhance everyday lives, but limitations and benefits are unknown, and the user perspective has been largely neglected. In particular there is a lack of empirical evidence to guide commissioners, developers and users.

1.8 Plan of Thesis

This introductory chapter explains the general perspective in which emotion in computing takes place. It illustrates some of the problems of definition of emotion in computing, describes the wider context and describes the general need for a more user-centred approach.

Chapter two gives an overview of the background literature relevant to emotions in computing, including the historical perspective of theories of emotion; terminology used about emotions and human communication and it surveys approaches to development of affective systems. It gives a detailed critique of methodologies for measuring and studying emotions in computing and existing knowledge about human computer interaction with affective systems.

Chapter three explains how my review of the literature resulted in a refined set of objectives, aims and proposed contributions of this work. It goes on to describe the theoretical and methodological basis behind the research and the research methods I decided to use.

Chapter four describes the main experimental hypotheses and procedure and use of a Wizard of Oz methodology to control an affective system, so that it appears to respond in a 'normal' human way to the users. Employing a two by two factor experimental design enables me to compare four groups of users interacting with an affective system that does or does not respond to their affective communication, and that they do or do not expect to do so. Later sections of the chapter describe the procedures used for analysis of the resulting data and the coding procedures.

Chapter five describes the experimental results. It reports how I compare the four groups of users and find significant gains in performance and satisfaction for users of affective systems and it identifies some interesting issues requiring further study.

Chapter six describes how using the same Wizard of Oz methodology, I experimented with two further groups of participants where the system appeared to respond to only particular affective expressions (identified as cues in my previous experiment). One group of users is trained to use specific emotional behaviours while the other group is not trained to do so. While not resulting in significant experimental results, the second study provides an important contribution through highlighting a number of problems for the potential use and evaluation of affect recognition technology and affective interaction.

Chapter seven discusses the findings and implications of the experimental work. Issues arising are discussed, including the complexity of understanding emotional expressions in context and the challenges of coding emotional behaviours and of validating coding procedures.

Chapter eight describes my conclusions and contribution. It presents a summary of the benefits and limitations of my work and the issues and opportunities arising from it for future work.

1.9 Summary

There are many issues to consider relating to use of emotion in HCI in the context of our culture. There are problems about definitions and technology driven development of systems, which do not take into account the complexity of human interaction. Work in this area so far is limited by the capabilities of prototype systems. Gold standard, empirical experimental studies as well as qualitative studies are needed in order to thoroughly investigate the user perspective of affective computing and to assess if, when and how users are willing to communicate emotion during human computer interaction and what the benefits or drawbacks might be. The following survey of the literature around these areas (in Chapter Two) informs the more detailed aims and objectives driving this work.

2. Affective Literature Review

2.1. Introduction

This chapter surveys the background literature relating to different perspectives on emotion, their uses in computing and their implications for the work described in this thesis.

First it shows how various theories, brain studies and models of emotion led to the focus of my experimental work (2.2). It explains how human communication science and studies of individuals in social contexts have influenced my study (2.3). It discusses the terminological challenges in this field and explains the terminology I adopt (2.4). It presents a brief survey of affective recognition and responsive systems (2.5) and the fit of emotion within human-computer interaction (2.6) that led to some of my decisions about objectives. The conclusions (2.7) summarise the implications arising from the affective literature review and their influence on my choices of methods. The resulting choice of methods is presented in the following chapter (Chapter Three).

2.2. Background Emotion Research

Emotion is a wide cross disciplinary research area with many different theoretical perspectives and models involved in it, which Scherer, a leader in the field, likens to 'a swamp?' (2004). There are several strands of work that are of particular interest in this context because they inspired decisions about my study.

2.2.1. Theoretical Perspectives

Emotion has been an important topic of study since prehistory with documentation dating back to the Greek philosophers and prevalent throughout most of the history of the study of psychology. Cornelius (1996, 2000) summarises the four main approaches in the last 125 years and how they converge:

1. Cognitive approaches are incorporated into the other three and these go back to Hellenistic philosophers who believed that thought and emotion are inseparable.
2. The Darwinian view (Darwin, 1872) is concerned with the *expression* of emotion, seeing this as an evolved phenomenon with important survival functions, seen similarly in all humans and related species. The work of Ekman (1982) and Plutchik (1980) on basic emotions follows this tradition, along with work to identify the neural architecture of emotions in humans and other species such as rats (e.g. Le Doux, 1996).
3. The Jamesian view follows Darwin – but is concerned with inner *experience* of emotion. Theorists suppose that bodily changes are perceived as emotion, so according to William James ‘ the nervous system of every living thing is but a bundle of predispositions to react in particular ways upon contact of particular features of the environment ……. we feel sorry because we cry, angry because we strike, afraid because we tremble’ (James, 1890 pp 449-50). James (1884) suggested that if we had no body we would be ‘excluded from the life of the affections’. The work of a number of theorists has been influenced by this view, for example theorists who look for appraisal patterns which link personality, physiological state, learning, history, context etc e.g. Fridja’s ‘action tendencies’ (1986). Hohmann went as far as proposing that people with impaired neurological organs, such as those with spinal cord injuries, might feel less emotion (Hohmann, 1966).
4. The Social Constuctivist perspective is the youngest, most diverse and controversial view, involving elements from cognitive, Jamesian and Darwinian work and drawing from anthropology and sociology. It is currently the main focus of attention in this field. Theorists see emotions as cultural products that owe coherence to learned social rules. Averill (1980) describes these as ‘social constructions’ that ‘can be fully understood only on a social level’. Subscribers to this view say that while the appraisal process may be a biological adaptation – appraisal content is cultural and cultural ‘scripts’ determine the form that emotion takes. For example, in a situation where someone kicks another and provokes anger, social constructivists see the angry response as due, not just to primitive responses, but to a cognitively dependent

> moral judgement based on social norms. The person who has been kicked makes a judgement that the kick breaks social norms of acceptable behaviour, appraises that it was not accidental and that feeling and expression of anger is justified in the form of retaliation or retribution.

There is very interesting work from all these fields feeding into the current theoretical perspective. For example Caroll Izard combines the first three approaches (1977), Smith and Lazarus (1993) take a mostly cognitive view but believe in universal facial expressions (like Darwinian theorists).

Implications: While acknowledging the importance of Cognitive, Darwinian and Jamesian views, my interest lies in human computer interaction that takes place in a social context and so I draw on the social constructivist approach in my work.

2.2.2. Studies of the Brain

In early studies of the brain and motor function it was thought that there was a direct relationship between certain areas of the brain and certain functions, but with the identification of 'association areas' (Luria 1973), that play a meta-management role, an appreciation of the complex nature of brain function was under way. In studies of affect, we are still learning about the exact nature of brain function in relation to emotional feeling, processing and expression. In neuro-cognitive studies the basis of language is gradually being revealed (Lamb, 1998; 2004).

The view of emotion as a physiologically based, complex system is supported by numerous brain imaging studies which have helped understanding of brain function and has led to the widespread acceptance that emotions result from a complex interaction between the amygdala (responsible for emotional experience), the hypothalamus (responsible for peripheral somatic and autonomic arousal) and the limbic system (responsible for the conscious experience). Neurological studies in animals suggest there may be 'fusion neurones' in the brain, that combine and recruit signals from different modalities. For example, in mammals perception of sound and vision is increased when two signals are received from the same source at the same time, and the smaller the two signals are, the higher the recruitment of them is, and the better the mammal perceives them (Rowland & Stein 2007). Following this line of thought, it may be that humans are better able to perceive changes in emotional expression if they are expressed by tiny changes and are expressed in a multimodal way, for example by sound and vision at once. In the context of emotional behaviours and affective computing, humans may recognise emotional behaviours that are very tiny and subtle and that are made in several different modes at once, much better than they recognise gross displays. For example a multimodal signal, such as a tiny smile made at the same time as a quietly spoken sound may be better perceived than a single mode signal, such as a big smile, or a loudly spoken word, made in isolation. At present imaging studies of humans are limited to those that can be carried out in laboratory settings or in the tunnel of a scanner. As scanning techniques progress and allow scanning of human brain activities to take place 'in the wild' in genuine social contexts, we will learn more.

To establish if and how physiological experience is linked to emotion, studies have been carried out with people with neuruological damage and it seems that in some regards they have less emotional experience. For example, Chwalisz, Diener & Gallagher (1988) describe spinal injury patients as reporting less intense emotions. They found that the higher the lesion, the greater the impairment for some emotions, although this was not true for love and compassion, where there was increased intensity which they postulated may be due to sensory memories from before injury and to facial feedback. Work with people with Parkinson's disease, (a motor movement disorder) which limits expression of emotion, found sufferers have less emotional awareness and lack emotional processing capacity (Wieser, Muhlberger, Alpers, Macht, Ellgring, and Pauli, 2006). Other facial feedback hypothesis studies found that enforced smiling led to people feeling happier (Bernstein, Clarke-Stewart, Penner, Roy & Wickens, 2000) and that preventing people with normal neurological functioning from smiling (by asking them to hold a pencil between upper lip and nose) resulted in reduced emotional experience of happiness (Davis and Palladino 2000). Studies suggest that feelings of love and compassion increase when appropriate facial expression is made (Hohmann 1966, Chwalisz, Diener et al 1988). Strack, Martin & Stepper found different postures associated with positive and negative affect (1988) and Fredrickson used cultivation of positive emotions for preventing and treating problems rooted in negative emotions, such as anxiety, depression, aggression and stress-related health problems, in order to optimise health and well-being (2000). If the emotions that people feel are dependent on physiology and if their physical behaviour influences the way they experience emotion, this has implications in human computer interaction. It leads us to question exactly what embodiment or physical movement or expression of emotion is afforded by a system and how that might influence how users feel while they are interacting with it.

There is a growing body of work examining the complex neural basis for emotions (LeDoux, 1996, 2003; Levenson, Ekman & Friesen, 1990). The limbic system, hypothalamus and temporal lobes and cerebral cortex of the brain are all involved in perceiving or reacting to emotional feelings. In evolutionary terms, reflexive emotional behaviours and reactions such as primal fear and aggression are associated with older, deeper parts of the brain and nervous system, whilst cognitive elements such as emotional management are associated with more recently evolved parts, such as the frontal lobes and cortex. Sensations can reach the amygdala directly from the thalamus or via the cortex. Robinson, Hoheisel, Windischberger, Habel, Lanzenberger and Moser in their review (2005) commented on the progress made using fMRI studies of emotions in the previous 8 years and some of the technical challenges with brain imaging of the limbic system in particular. Methods are rapidly progressing, for example, with very recent work using multivariate pattern analysis in order to understand overall spatial patterns of activity in the auditory cortex rather than just charting activity in specific regions. For example, Ethofer presented people with pseudowords spoken in five ways - with anger, sadness, relief, joy, or no emotion – during fMRI of voice sensitive cortices (Ethofer, Van De Ville, Scherer, & Vuilleumier 2009).

Some regions of the brain have been identified as associated with particular emotional functioning. Although lateralisation is evident, 'the laterality of emotions is far from determined' (Achuff, 2001) and this is an interesting area to consider. As early as 1881, Jules Bernard Luys identified that those with a stroke affecting the left hemisphere reported emotional lability and volatility (Parent, 2002). Current research suggests both hemispheres are important, in different ways. The proponents of the right hemisphere hypothesis hold that the right hemisphere is more dominant for all forms of emotional expressions and perception (e.g. Tucker, Watson & Heilman, 1977; Hellige, 1993; Borod, Andelman, Obler, Tweedy & Welkowitz, 1992). In contrast, the valence hypothesis suggests that hemispheric asymmetry depends on valence of emotion, with the right hemisphere more concerned with negative emotions and the left with positive emotions (e.g. Sackheim, Gur & Saucy, 1978; Davidson & Fox, 1982; Fox 1991; Davidson, 1992). If one were to combine these two theories, it would suggest that the right hemisphere is more involved in perception and emotional experiences and the level of activity determines if the experience is positive or negative (with higher activity levels relating to negative affect), while the left hemisphere is more involved in the process of expressing emotions. Future neurological research may clarify this. New techniques such as facial Electromyography (EMG), Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI) and event related potential (ERP) are rapidly increasing our understanding of the neurological basis of emotions, for example using EMG, Dimber and Petterson (2000), found increased activity levels for muscles on the left side of the face to visually presented emotional signals, supporting the right hemisphere hypothesis; Reiman and his associates (Reiman, Lane, Ahern, Schwartz, Davidson, Friston, Yun & Chen, 1997) demonstrated that different parts of the brain and different neural systems are associated with different aspects of emotion and that some areas are involved with management of emotion, rather than with discrete emotions; Cunningham, Raye and Johnson (2004) have used fMRI to demonstrate the complex circuitries of neural mappings, and Ochsner, Bunge, Gross & Gabrieli (2002) demonstrated areas used during the cognitive regulation of emotion.

We are learning more about various aspects of emotion such as valence and appraisal and of the relationship of emotion as a human system to other complex high level human systems such as cognition or stress (Lazarus & Folkman, 1984). Scherer (1995) and Lazarus (1991) emphasise the importance of higher order cognitive processes, such as using working memory and appraisal. Emotional events are typically distinctive events, and distinctive events are easily retrieved from memory. This leads to good memory for the “gist” of emotional and consequential events, but strong emotion appears to disrupt memory processing, leading to a narrowing of processing and this leads to loss of memory for peripheral details. Vivid 'flashbulb memories' (Brown & Kulik, 1977) can be formed for emotion related events. Many researchers have found improved memory for emotionally arousing or emotionally related events (e.g. Christianson & Loftus, 1990; Rubin & Kozin, 1984; Revelle & Loftus, 1992; Schacter, 1996). Richards and Gross (2000) suggest that the process is complex, with individuals that use a high amount of emotional regulation (the 'stiff upper lip' syndrome) having reduced memory of events. In order to adapt to aversive events, humans have adopted many strategies. One described is "reappraisal", where the negative affect associated with an aversive event is reduced due to a cognitive transformation, for example someone seeing a scary face might decide to focus on funny aspects of the face, in order to reduce negative affect. It is expected that long term studies will provide more data on emotion regulation and cognition (e.g. Wolfe & Bell, 2004).

Implications: In time this wealth of work may lead to better conceptualisation and labelling of emotion. Meanwhile it tells us that emotion is immensely complex, that our knowledge of emotional mechanisms is far from complete with much still to be established and that we should remain open to wide understanding of both felt and expressed emotion.

2.2.3. Models of Emotion

Theories, along with brain studies, have led to new models of emotion, which help to visualise the complex interactions involved in inner feeling and outer expression of emotion. Early signal processing models of communication were developed in World War Two to help analyse communication of artillery issues. As theories of emotion have developed, these early models of communication provided the basis for subsequent elaborations and led to current more complex models of emotion, often involving different levels of emotional functioning, a number of which have influenced the field of emotion in computing and are commonly used in affective systems. Four broad approaches have influenced models.

Dimensional models. One approach in models of emotion is based on the premise that any emotion can be described in terms of certain dimensions, although the dimensions differ from one model to another.

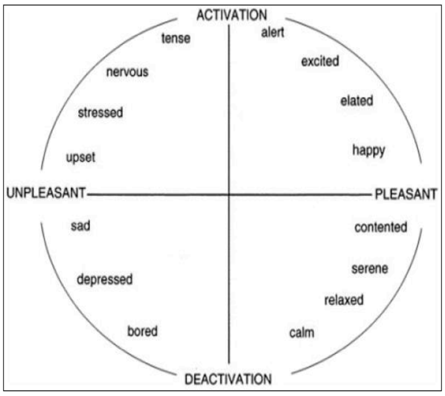


Figure 2. Russell's circumplex

Russell's circumplex model of affect (see Figure 2) was developed on the basis that affective states arise from cognitive interpretations of core neural sensations that are the product of two independent neuro-physiological systems. Russell (1980, 2003) originally plotted constructs of emotions (determined by factor analysis) against two dimensions –activation (the degree of intensity or arousal of an emotion) and pleasantness (the valence - how positive or negative the emotion is) in a circular representation. Refinements of this model include use of three dimensions: affiliation (communion, warmth, evaluation, etc.), power (agency, dominance, competence, potency, etc.), and activation (arousal, action readiness, affect intensity, etc.). Watson and Tellegan changed the orientation and proposed four dimensions: pleasantness, engagement, positive and negative affect (1985).

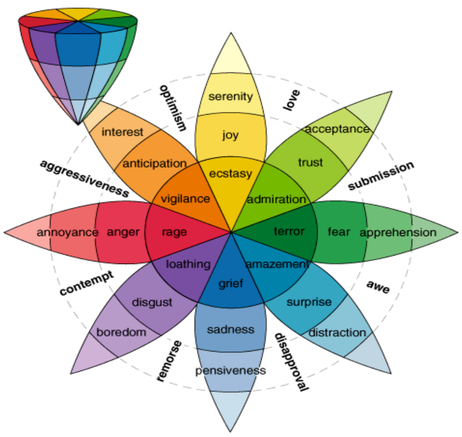


Figure 3. Plutchik's cone (1980).

There are criticisms of two-dimensional models as being too simplistic (e.g. Fontaine, Scherer, Roesch & Ellsworth, 2007). Plutchik (1980, 2003) proposed a cone shaped model (see Figure 3) with intensity of emotional experience represented by depth, similarity by nearness and four pairs of opposites, all represented by colour coded segments.

Theorist	Basic Emotions	Basis for Inclusion
Arnold	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness	Relation to action tendencies
Ekman, Friesen, and Ellsworth	Anger, disgust, fear, joy, sadness, surprise	Universal facial expressions
Frijda	Desire, happiness, interest, surprise, wonder, sorrow	Forms of action readiness
Gray	Rage and terror, anxiety, joy	Hardwired
Izard	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise	Hardwired
James	Fear, grief, love, rage	Bodily involvement
McDougall	Anger, disgust, elation, fear, subjection, tender-emotion, wonder	Relation to instincts
Mowrer	Pain, pleasure	Unlearned emotional states
Oatley Johnson-Laird	Anger, disgust, anxiety, happiness, sadness	No propositional content
Panksepp	Expectancy, fear, rage, panic	Hardwired
Plutchik	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise	Relation to adaptive biological processes
Tomkins	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise	Density of neural firing
Watson	Fear, love, rage	Hardwired
Weiner and Graham	Happiness, sadness	Attribution independent

Table 2. Ortony and Turner, (1990) 'What's basic about basic emotions?'

Basic emotions A second group of models are based on the concept that there are basic emotions that serve evolutionary functions, have a neurological basis and that cannot be reduced to more primitive components. Most emotion experts think there are a few 'basic emotions', although they don't all agree what they are (as shown in Table 2.) or why they are 'basic' (Ortony & Turner 1990). Paul Ekman (1982, 2003) has been very influential with his studies of facial expressions. He originally suggested six 'basic' emotions, which he claimed to be universal, and that mapped to certain autonomic activities, such as increased heart rate mapped with anger. Ekman's original basic emotions model supposed that a discrete and independent neural system subserve every emotion.

A frequent criticism of Ekmans's work is that he relied on static rather than dynamic portrayals of emotion. His work is often somewhat misrepresented, for example some researchers interpret the concept of 'universal' expressions to mean that they can reliably take any example of a particular facial expression to represent an expression of a particular basic emotion, (e.g. that a smile is always an expression of joy, when in fact it might be a social signal such as a greeting). In fact Ekman says that interpretation should always depend on context. This is important to my work because it underlines the importance of social context, which has been neglected in some studies.

Ekman has now developed his description of what constitutes a basic emotion. He describes different possible meanings of 'basic' emotions. These range from to descriptions of families of basic emotions and discussions of emotional terms that might be included as emotional states or might be cognitive states, although he expresses doubt about some categories such as 'guilt'.

> "From this perspective, fear, anger, disgust, sadness and contempt, all negative emotions, differ in their appraisal, antecedent events, probable behavioral response, physiology and other characteristics .... So, too, amusement, pride in achievement, satisfaction, relief and contentment, all positive emotions, differ from each other. This basic emotions perspective is in contrast to those who treat emotions as fundamentally the same, differing only in terms of intensity or pleasantness" (Ekman, 1999 p.45).

Cognitive models Sloman's Cogaff model (2001) (see Figure 4) describes levels of emotional reactions in evolutionary terms as reactive mechanisms, deliberative reasoning and meta-level management (reflective processes that act on perception, central processing and action). He describes different classes of emotions that relate to these levels. Primary emotions (the oldest in evolutionary terms) that link to reactivity, secondary emotions relating to the deliberative level mechanisms and tertiary emotions relating to meta-management issues such as loss of attention. This model has considerable flexibility of application as by changing the size of each level it is possible to apply the model to very different situations, for example Sloman suggests that an insect such as a house fly might operate on only the reactive mechanisms level.

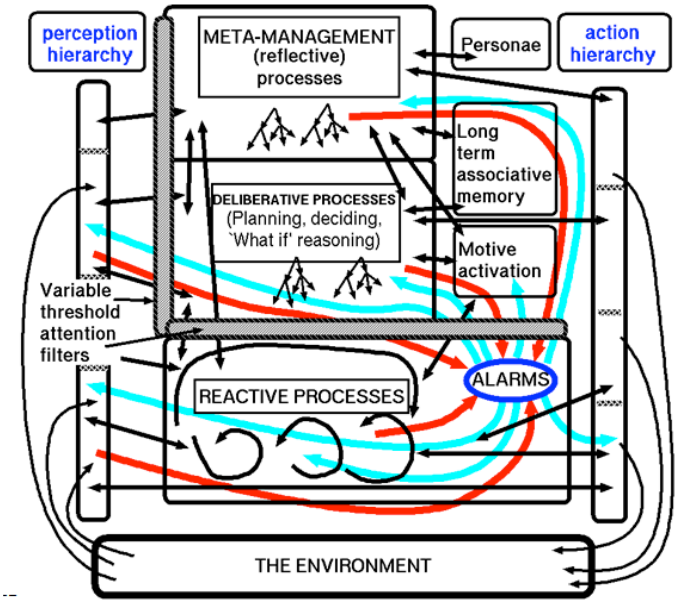


Figure 4. The Cogaff model (from Sloman's Cognition and Affect Project website)

Minsky (2006), from a strong background in artificial intelligence describes the brain as an emotional machine, with six high functioning structural levels of operation (see Figure 5).

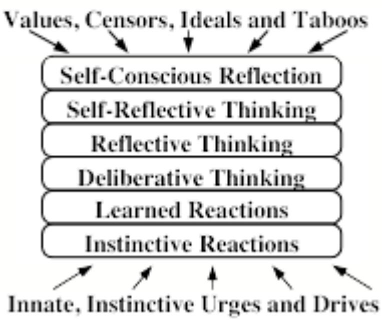


Figure 5. Minsky's six structural levels of emotion from 'The Emotion Machine'(2006).

Ortony, Clore and Collins (OCC) model (1988), designed as a computationally tractable model, (see Figure 6) is often used as a basis for development of affective applications, and particularly for that of embodied characters or agents. It considers valenced reactions to consequences of events, actions of agents and aspects of objects. Although it is possible to consider social interaction within this model, it is still lacking in some respects as application developers still have to decide on representation of the emotional state of the agent (Bartneck 2002). This is an interesting model often used in building affective systems, for example Ortony et al's (1988) 'OCC' model was built to facilitate implementation in a computerised system:

> ".... we would like to lay the foundation for a computationally tractable model of emotion. In other words, we would like an account of emotion that could in principle be used in an Artificial Intelligence (AI) system that would, for example, be able to reason about emotion."
>
> (Ortony, Clore and Collins, 1988, p. 2)

The OCC model is widely used in artificial intelligence (AI) systems such as embodied conversational agents (ECA's) (e.g. Bartneck, 2002).

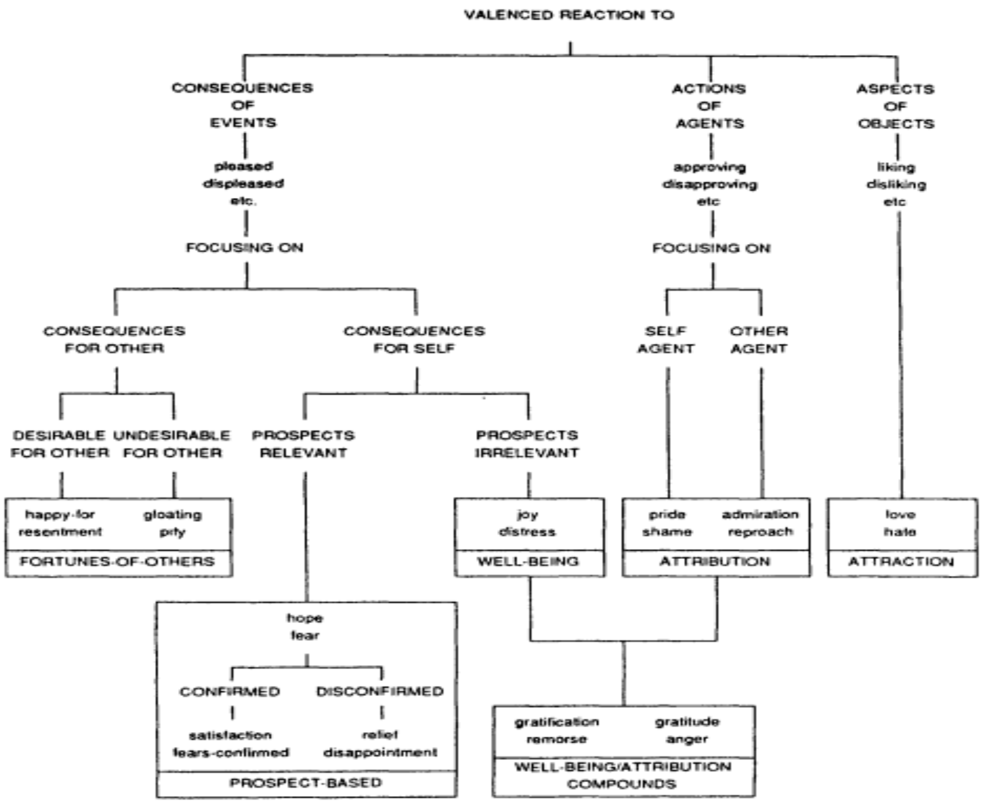


Figure 6. OCC Model (from Ortony Clore & Collins, (1988), The Cognitive Structure of Emotions pp19)

Interpretation of emotions is a particularly challenging area. Scherer (1987, 1999) developed a component process model. In this model (Figure 7) emotion is seen to consist of biological, neurological and psychological systems operating on a number of interacting mental and physiological processes (e.g. visceral, behavioural and reflective processes) using an invariant sequence of stimulus evaluation checks. The checks are a result of appraisal – i.e. as an organism constantly assessing the state of the environment in relation to goals, well being etc. with subsystems responsible for a range of mental expressive and physiological responses.

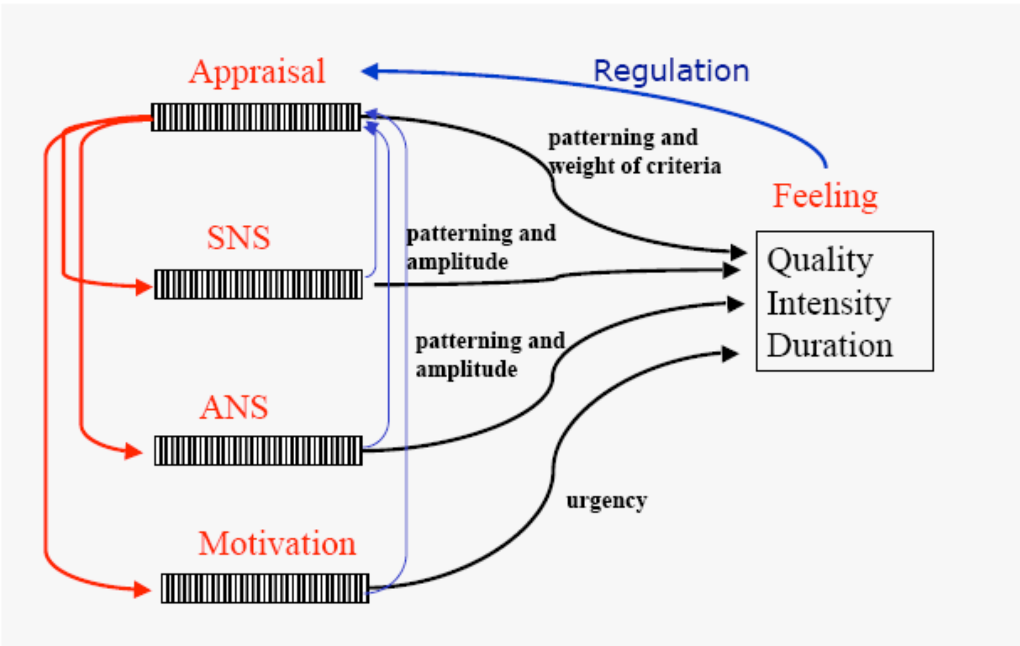


Figure 7. Scherer's component process model from Humaine website

Norman (2004) in his book about emotional design of products describes applying these levels of emotional processing to design (see Figure 8). He describes biological/visceral (e.g. gut reactions), neurological/ behavioural (e.g. response to the pleasure and effectiveness of a product) and psychological/cognitive processing (e.g. longer term reactions linked to memories such as pride).

> "….three different aspects of design: visceral, behavioural and reflective. Visceral design concerns itself with appearances……Behavioural design has to do with the pleasure and effectiveness of use……..Finally reflective design considers the rationalisation and intellectualization of a product." (Norman, 2004 p.5).

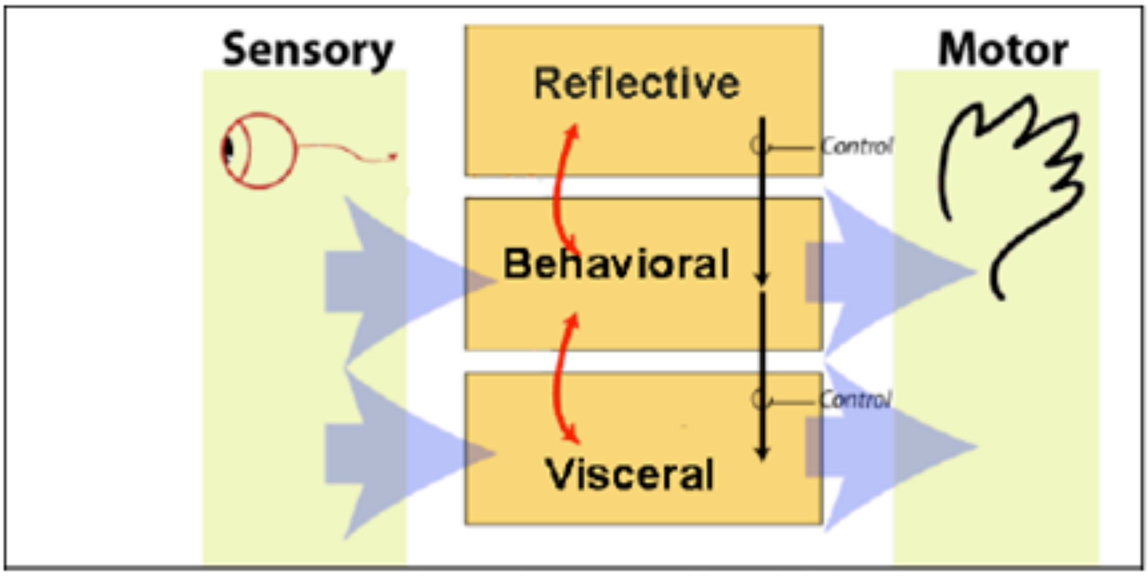


Figure 8. Don Norman's levels of emotional processing (from 'Emotional Design' 2004)

Interactional models In their paper 'Affect: from information to interaction', Boehner, DePaula, Dourish & Sengers (2005, 2007) proposed an interactional model of emotion, that seeks to depart from the information processing models in order to satisfy the complexities and criticisms of other models that arise in HCI. They see emotions as interactional products rather than informational objects, that move and change during interaction, that are complex, and that can be incomplete or ambiguous:

> "Drawing on cultural, social, and interactional critiques of cognition which have arisen in HCI, we introduce and explore an alternative model of emotion as interaction" (p.59).

Boehner, DePaula et al view emotion not as discrete units that can be transmitted and received, but as a social practice, with complex felt emotions that develop during human interaction and that may or may not be transmitted. They emphasise the importance of *social* context:

> "emotion as interaction: dynamic, culturally mediated, and socially constructed and experienced";
> "everyday action as situated in social and cultural contexts that give them meaning".

And they suggest a view of future systems with:

> "new goals for the design and evaluation of affective systems - instead of sensing and transmitting emotion, systems should support human users in understanding, interpreting, and experiencing emotion in its full complexity and ambiguity."

Study in particular domains has led to specific models and refinements, for example Fredrickson (1998) 'broaden and build' model was developed to understand resilience and understand how people use the power of positive emotions for well being, for example after the 9/11 terrorist attacks in the USA. She takes an evolutionary view of how emotions function and finds that negative emotions such as fear, anger, and sadness narrow momentum to specific actions that are likely to promote survival, while positive emotions such as loving kindness, joy, interest, and contentment lead individuals to broaden and build resources – also to promote survival (Fredrickson 1998; Cohn, Fredrickson, Brown, Mikels & Conway, 2009).

Implications: It is important to consider models of emotion and how these have informed and are informing development of systems. Dimensional models have proved very useful in the measurement and coding of emotion related behaviours. 'Basic' emotional expressions have been very influential in early human-computer interaction studies. Cognitive models and systemic theories are rich and very complex and help us to understand the complexity of emotion. Computational models of affect are being developed that are proving very useful, particularly in the field of artificial intelligence, although they do not represent the full complexity of human emotional interaction. They are useful because they provide a framework that programmers and developers can use to map aspects of emotion to aspects of the system and different models have different pros and cons (Bosse, Gratch, Hoorn, Pontier & Siddiqui, 2010). Interactional models are very relevant for this thesis. My approach fits very well to the Interactional Model (Boehner, DePaula, Dourish & Sengers, 2005) as I consider the context and quality of the interaction that takes place. Concepts of levels of emotional operation and of appraisal are relevant in the practical real world task that I go on to develop.

2.3. Human Communication and Emotion

2.3.1. Multimodal Human Communication

Human communication consists of verbal communication, which can be modified by paralinguistic features such as tone of voice, and of non-verbal communication. Communication signals can be transmitted simultaneously or sequentially using multi-sensory *modes* including:

- audible sound; such as tone of voice, voice quality, intonation pattern, rate of speech, choice of words, accent, dialect, etc;
- visual appearance; such as choice of clothing, accessories, environment etc;
- visual positioning; such as personal space, proximity or distance, posture, etc;
- visual use of body movements: such as gestures, signs, facial expressions, tensions etc;
- touch; for example pressure exerted during a handshake etc;
- taste and smell; such as sweatiness, pheromones or use of perfume (Chen & Haviland-Jones, 2000).

These sensory modes are all modified by context and time. So for example, close physical proximity might generate a positive experience between two people who know each other well. It might be socially acceptable, despite discomfort, on a crowded tube train. But close proximity by a total stranger, when not imperative because of spatial constraint, is likely to feel uncomfortable or even threatening. If we see someone standing momentarily on a street corner, we might interpret his behaviour and emotional experience in one way, but if we observe the same person standing there for several hours, we might place a different interpretation. We follow many unspoken rules of expression and interpretation and about acceptable use of various modes. Timing and context of expressions add another layer that affects both our understanding and expression of emotional states. The interplay of modes, timing and context might be idiosyncratic, or socially or culturally determined.

In spoken language, paralinguistic features such as tone of voice intonation patterns etc. add to the meaning of words. Other non-verbal means of communication such as body posture, positioning, gestures, facial expression and gaze give further meanings. Even the way we dress, how we present ourselves visually and where we place ourselves in the world and surround ourselves with artefacts, all communicates something about us and about our emotional state. From a 'flashbulb' smile or graphic gesture, to the tiniest change in posture or vocal tension, from the cradle to the grave, both consciously and unconsciously we constantly transmit and monitor multimodal expressions.

In written language we convey paralinguistic features by variables such as our choice of media, style of writing, choice of words (e.g. shared codes) and use of graphics e.g. use of capitalisation and symbols such as emoticons, etc.

As well as by using all the various modes, meaning can be changed by subtle changes in the timing of multimodal behaviours or events. All multimodal expressions are modified by their appearance and position in time, space and context. Timing can alter any given expression to look different or mean something different, to different people. A multimodal expression might change in meaning if it is sustained for a period. For example a sustained look of anger might be interpreted as more intense than a fleeting one. Depending on the selection and synchronicity of different modes, the exact meaning of a multimodal expression may be altered, for example timing an inflection on part of a word or sentence, or using a gesture at a certain point in the conversation. We often use mixed messages in everyday interaction, for example a smile, given along with spoken words, can alter the meaning of a harsh word, or show irony.

Different modes of communication may be more relevant in different situations. Information from different modes may or may not be additive. It may be that particular modes or combinations of modes are significant for particular types of message or may even map more exactly to underlying emotional states. It is possible that certain modes are more important to convey certain types of information. Certain modes may attract more attention than others, for example our cerebral cortex has far greater proportional representation of the face and hands than other areas, (see Figure 9.), so attention to facial expressions or hand gestures may dominate.

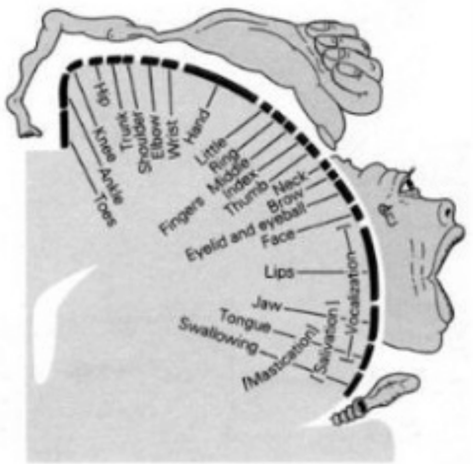


Figure 9. diagram of the motor homunculus (from Penfield & Rasmussen, 1950).

Alternatively we might attend to particular types of information regardless of the mode used to convey it. For example information about cognitive and emotional states might be important to us in all or in certain contexts and so attract most attention, regardless of the mode used for conveyance. It may be that, as has been found in other mammals (Rowland & Stein, 2007), that we attend more to smaller stimuli in certain combinations than to discrete signals, whatever their amplification.

Our communicative system has built in redundancy – information is often conveyed by more than one means at once. This may have an evolutionary purpose to ensure that messages are more likely to be noticed and understood, and to suit different preferences of communication partners. Communication is an imperfect art, and even in favourable conditions messages can be missed or misunderstood. For example, one estimate is that in human-human communication, only 55% of non-verbal communication is accurately detected (Mc Gilloway, Cowie, Douglas-Cowie, Gielen, Westerdijk, & Stroeve 2000).

Some modes of communication may be more prone to conscious or unconscious control and the extent to which emotions are revealed or hidden varies between people and between situations. Ekman (2003) describes 'leakage' of underlying emotions during facial actions but this is contended by Bavelas (Bavelas, Black, Chovil & Mullett, 1990).

We use multi-sensory displays to augment or add emphasis to what is said, to alter meaning, or sometimes instead of verbal language. This adds greatly to the richness of experience of communication, and to the effective exchange of information that can be achieved. For example we use non-verbal signals to convey factual information, reveal affective states, show our communicative intent and control communication flow. We may use them when alone or with an imagined audience as well as in interactive situations. Bartneck and Lyons (2009) review 'face based' HCI, including computational methods for analysis of facial information, modeling of internal emotional states, and methods for graphical synthesis of faces and facial expressions in artificial intelligence (AI) applications. They cite Schulz (1981) who suggests messages consists of: facts, relationship, appeal and self-revelation.

The study of facial expression, particularly by Ekman, has led other research into non-verbal communication in affective computing. Ekman's work on facial expressions has been hugely influential and very important in driving forward the field of affect, but there are some misinterpretations and criticisms of it. His original work on basic emotions, (and 'universal' expressions) have led to many affective applications that only consider 6 basic expressions, assuming them to be all encompassing.

Ekman (2003) differentiates between involuntary *emotional* expressions (for example a smile that reveals a happy inner state) and intentional, voluntary *communicative* facial expressions (for example a smile to signal greeting). These can be distinguished from emotional expressions if they:

- are part of the conversation structure and flow of talk. Rule governed emotional expressions can occur while talking but are not part of structure, but rather part of semantics.
- show limited scope of movements deployed. Conversational signals are single movements in one region of face e.g. pressed lips or raised eyebrow, contrasting with emotions that run or sweep across the ace (though these maybe limited if consciously inhibited.)
- use movements that are easy to make. Ekman notes that motion sweeps are hard to imitate.

Emotional facial expressions can occur even when there is no conversational partner, although they may be managed (attenuated, amplified, inhibited or covered with another sign) or partly concealed to comply with cultural social rules. Ekman acknowledges that emotional expressions are only one subset of facial expressions used in various domains. Facial expressions can also convey involuntary messages, such as burping. Krause, Steimer-Krause & Ulrich (1992) suggests emotional role-playing or re-enactment of emotions can involve use of real emotional expressions.

Communicative expressions are voluntary and carry semantic meaning in conversational situations. They tend to use a limited scope of movement and only part of the face, and are physically easy to initiate, including actions such as kissing; symbolic gestures or emblems, such as winking; conversational signals, such as raising an eyebrow to emphasise a word; false expressions; referential expressions when referring to an emotional state such as transformation by pulling mouth back while saying the word 'afraid'. Ekman recognises a whole range of functions of verbal expressions and advised that facial expressions must always be interpreted *in context*, (which should include the environmental, social and verbal context) but this caveat was often ignored in early affective system research.

The face, in particular, is a multi signal device and according to Ekman (1997) facial movements may be:

- actions (eating, kissing etc);
- symbolic gestures (e.g. winking) (emblems in Ekman's terminology) or choosing to say a word as a symbol;
- conversational signals – to illustrate speech e.g. raised eyebrow to add emphasis while speaking – this can be used in a voluntary or involuntary way to emphasise a word;
- false expressions – these closely resemble real emotions and are used as deceit. Often real and false expressions can be identified, for instance real enjoyment can usually be distinguished from false expressions of it;
- referential expressions – these are used during conversation when referring to an emotional state. They are transformations of emotional expressions. You see part of whole expression only, e.g. mouth pulled back on word 'afraid'.

These occur more quickly if involuntary, but may be extended to last longer if used voluntarily, or if used in mock reference (e.g. a humorous reference). They are distinguished from real expressions so as not to confuse others and so as not to have to re-experience emotion;

- emotional role-playing - Rainer Krause (Krause, Steimer-Krause & Ulrich, 1992) suggests another type seen in psychotherapy sessions – this involves enacting another's emotions – like referential use or momentarily becoming emotional and then probably showing actual emotions.

Emotional awareness is so basic to human life that even a glance at another person leads to assumptions about personality, mood, emotional state etc. During interaction, however superficial, we are tuned to the slightest emotional clues. Only a few interactive behaviours or emotional expressions can be reliably mapped to underlying emotions (e.g. Ekman (1982) has identified six static facial expressions that when displayed he believes to have universally understood meaning).

Researchers in the field of human communication science give compelling arguments for the *non emotional* roles that facial expressions and other forms of non verbal communication such as gesture, tone of voice etc play during interaction. It is unlikely that people intend all their behaviours to primarily convey emotions and Bavelas argues strongly that people convey only those non verbal messages that they intend to convey during interaction (Bavelas, Black, Chovil & Mullett, 1990) and that there are different conventions for speakers and listeners. Ward and Marsden (2004) suggest intentional affective communication may be easier to detect than underlying feelings.

There are commonly held beliefs about the meanings of some non verbal behaviours, which have little or no scientific basis or are based on misinterpretations of work. For example it is often said that folding the arms is putting a symbolic emotional barrier between speakers. Roberts & Bavelas (1996) point out that this could equally convey that the speaker is physically feeling cold. Bavelas has shown that many common interpretations of non-verbal behaviours are based on no more than popular fiction and she proposes that gestures may have many functions, depending on context, for example they may be primarily used to replace, accentuate or add to the meaning of a spoken communicative episode. So sure is she of her stance that she has offered $1,000 to anyone who can prove that they can reliably detect an emotion in a subject, which that subject did not intentionally communicate.

Implications: We need to understand how humans communicate in general and how and when they communicate emotions. In the affective computing field early studies relied heavily on facial expression of basic emotions as described by Ekman and often overlooked alternative reasons for non-verbal communication. We should seek to go beyond 'basic' emotions and consider multi-modal aspects of communication; understanding of multi-modal behaviours in their social context; and understanding of when such behaviours are primarily intended for communication of factual or emotional information, or both.

2.3.2. Variables

We have established that communication of emotion is both a continuous and usually an intentional signal with a number of individual differences to be considered, as well as interactional, social and cultural variables.

1. **Environmental constraints** and perspective can alter perceptions of emotional expressions. For example Noh masks (used in traditional Japanese theatre) are tilted by tipping the head to display different angles of the mask to the audience. The mouth appears up-turned (smiling) when viewed from above, or down-turned if viewed from below (see website for Lyons, Campbell, Plante, Coleman, Kamachi & Akamatsu, 2000).

2. **Temporal and contextual differences** influence affect. Familiarity means that it is easier to interpret the expressions of someone you know well, partly because you know their usual usage of expressions and will recognise and respond more sensitively to variations from their norm. For example someone who frequently has a 'smiley' expression will have to give a much broader smile to make it significant in comparison to someone with a habitually sad face. Temporal changes, such as mood on a particular day or at a particular time may also lead to changes in emotional

perception or displays. Even when we say nothing and do nothing, then that very absence of communication is sending a message, especially if such communicative silence is unusual for us as an individual or inappropriate for the communicative setting. In certain situations where a facial display is expected a blank face can give a very powerful message, For example, Givens (2009) describes someone showing a blank expression rather than laughing in response to a joke.

During interaction, however superficial, we are tuned to the slightest emotional clues and make judgements about personality, mood, emotional state etc. Even small favours and kindnesses such as finding a small coin in the street have been found to induce a positive affect in people who in turn are likely to 'pay it on' and induce positive affect in others (Isen, 2000).

We should consider discourse as a series of acts that match to goals (Searl, 1969; Austin, 1962) and should consider how they fit within social rules such as politeness (Brown & Levinson, 1987). Even the way we laugh is rule governed (Provine, 2000).

Contextual differences can make any expression look different or mean something different. The expression described by Darwin as typical of 'grief' is the one Woody Allen uses to signal delivery of a joke (Bakeman & Gottman, 1997).

3. **Physiological differences** may mean that someone with a wider mouth or stronger cheek muscles may appear to smile more than an individual with other physical characteristics. Similarly changes in neurological functioning and chemical balance may lead to different affect traits or instances of expression. For example following a stroke or brain injury, muscular weakness commonly impairs expression on one side of the face and may also cause emotional lability, where individuals sometimes show an inappropriate, uncontrolled, decreased or increased degree of expression of their emotional states.

4. **Susceptibility to affect** has been found to vary – Zelenski and Larsen (1991) identified factors from personality tests relating to susceptibility. Some people are more susceptible to positive or negative affect than others.

5. **Personality differences** may explain some of the differences in susceptibility to various kinds of affect. Psychologists have identified links between personality traits and affect. For example, Tellegen (1985) links positive affect with extraversion and negative affectivity with neuroticism. Different people have different motivations that lead them to behave in different ways (Oulasvirta & Blom, 2008).

6. **Emotional management skills** vary and emotional awareness and expression are a fundamental human ability, important in personal and social life (Izard, Kagan, & Zajonc, 1984), and very complex (Young, 1996). Goleman (1995) has shown how management of emotions and their display or 'emotional intelligence' is important and suggested that an emotional quotient (EQ) can be assessed. The term "emotional intelligence" has been coined to describe attributes and skills related to this concept (Koonce, 1996). This links to the 'emotional labour', first described by Hochschild (1983) in relation to the way airhostesses were expected to manage their feelings and displays of feelings by non-verbal signals such as smiling, regardless of their inner feelings. This idea has gained momentum and is now widely studied in diverse domains including health, leisure and the home (e.g. Smith, 1992; Rojek, 2009). Cole, Martin & Dennis (2004) have explored how emotion regulation and management develops in children and changes with age.

7. **Perceptual or cognitive capacities** may be limited in some individuals so that they cannot recognise any or all aspects of a multimodal signal. For example a blind person will not see the facial expression of others, or individuals with autistic spectrum disorders have particular difficulties with perceiving, interpreting and displaying emotional signals and in following social and emotional display rules. There are some intriguing examples of individuals who have impaired or missing ability for facial recognition and perception, and yet still have awareness of emotional expressions. For example some individuals have 'face blindness' (prosopagnosia) (Grüter, Grüter & Carbon, 2008).

8. Sex and gender differences Sex and gender identity influence the generation, experience and display of emotions. There are both differences and similarities between men and women, found in neural, experimental, behavioural and physiological correlates of emotion (e.g. Tannen, 1990; MacGeorge Gillihan**,** Samter & Clark, 2004; Schienle Walter, Schäfer, Stark, & Vaitl, 2005; Baron-Cohen, Knickmeyer & Belmonte, 2005). It is difficult to establish if differences are in feeling or perception, response or regulation (Wager & Oschner, 2005). These may vary with the type of emotion, for example Schienle, Walter, Schafer, Stark & Vaitl (2005) found greater neurological responses to violent or aggressive scenes in men, than in women. Men and women use different communication strategies (Tannen, 1990). Females are born with an increased tendency to attend to emotional facial displays (Baron-Cohen, 2003), and baby girls make more eye contact with care givers (Hittelman and Dickes, 1979), although these differences may be learnt very early in life to meet the expectations of cultural sterotypes. Leeb & Rejskind (2004) found evidence that mutual eye gaze behaviour of newborns was differentiated by sex within weeks of birth. Women have been found to be more expressive for most emotions, with the exception of anger (Hall, 1984). Women and men do not differ in their physiological responses to some emotional stimuli (Casa, Hofer, Weiner & Feldon, 1998), though the picture is not clear and other studies show enhanced responses to startle stimuli in women. Women express more negative emotions in intimate relationships, whilst men show more reactivity when viewing a hostile facial expression (Brody, 1999). Women show more electromyogram changes in response to facial expressions (Dimberg and Lunquist, 1990). Winton, Putnam & Krauss (1991) found that men and women have some different autonomic physiological responses to affect. Women have been found to be more facially expressive and less able to inhibit facial expressions when asked to (Grossman and Wood, 1993; Wagner Buck & Winterbotham, 1993). Barret, Robin, Pietromonaco & Eyssell (1998) found that there were differences in reports on amount of emotion felt, when people were interacting with the opposite sex, although they found that differences in emotions reported retrospectively, disappeared when considered on a moment to moment basis. Males are more likely to match emotional expressions to a self-presentation goal (Levine and Feldman, 1997).

9. **Age** also affects emotional experiences and their recognition and expression. Emotional capacity seems inborn with newborn infants attending more to real faces than composite faces and able to imitate facial expressions such as tongue thrusts within hours of birth (Nadel, 2004). Newborns smile in their sleep, and even anencephalic neonates (who are born without large portions of the brain, including those associated with control of emotion) show patterned facial responses to certain stimuli, such as disgust expressions to bitter-tasting substances and smiles to sweet tastes (Steiner, 1973). Studies of blind infants and children generally support the position that many facial expressions result from innate factors rather than depending on visual learning (Charlesworth & Kreutzer, 1973). Cole, Jenkins & Shott (1989) describe how blind and sighted children control their expressions. Innate reactions such as interest, disgust, physical distress and startle response, lead to the capacity for the primary emotions, anger, surprise, joy, and sadness, which are usually evident by four months of age, with fear added between months five and seven. Predisposition to temperament traits may also be inherited or present from birth, with functional organisation of emotions developing in childhood and beyond (Hiatt 1978; Chess & Thomas, 1996) linked emotional development to the development of social networks. By one year old children are using their mother's emotional expressions as a guide to what is safe or desirable (Hiatt, 1978; Hornik, Risenhoover & Gunnar, 1987). Children, as young as 12 months old, appear able to read and respond to TV actor's emotions and they respond particularly to aggression (Mumme and Fernald, 2003). With self-awareness at about 24 months come secondary emotions of shame, envy, pride, guilt, contempt and embarrassment, and early attempts at emotional regulation compliance.

Emotion regulation in relation to age has become a central topic in child development (Eisenberg 2000). As language, cognition and social awareness develop children use these to regulate their emotions, for example Cole, Martin & Dennis (2004) report that preschool children hid their disappointment more when with a researcher than when alone. It is suggested that by school age emotions and cognition are integrated (Bell and Wolfe, 2004). There is evidence that children are more influenced by emotions and less rational than adults (Belavkin, 2001). Social learning and maternal attachment may influence emotional development, and according to Vygotsky's social development theory (1978) this is dependant upon those within their zone of proximal development and is a complex process and not related to set stages. By teenage years, peer groups are the most powerful social referent, and there is evidence that in Western cultures, adolescents perform poorly on emotional processing tasks during their teenage years and that adolescents pass through a phase where their ability to quickly recognise other people's emotions nosedives (McGivern, Andersen, Byrd, Mutter, & Reilly, 2002). These findings are supported by neurological research that shows development and activation of brain areas associated with emotions and cognition changes in childhood and particularly at puberty and again in later life (Bell & Wolfe, 2004; Williams & Drolet, 2005).

Sociocognitive theories propose that with age there is increased ability to regulate emotions due to greater life experience of emotion control strategies, for example older adults use more passive emotion focused strategies, such as hiding their feelings when handling everyday problems (Blanchard-Fields, Stein &Watson, 2004). Older adults are more effective at avoiding negative affect and using emotion regulation techniques for decision-making (Charles & Cartensen, 2007). They also remember negative events less than positive events (Charles, Mather & Cartensen, 2003). Neuro-imaging techniques show that older individuals rely more on prefrontal (especially left prefrontal) networks, rather than on amygdala activity, suggesting more use of emotional reasoning in responses (e.g. St. Jacques, Bessette-Symons & Scabeza, 2009).

10. Social positioning Rules about use of multimodal expressions vary depending on social status and hierarchy. Social positioning dictates that the rules about use of multimodal expressions vary depending on social status and hierarchy, for example a student might greet a teacher or older person more formally than a peer, or choose to sit in a different position in relation to them and those in a position of power (such as teachers) are more likely to 'catch' the emotions of those they supervise (Hsee, Hatfield, Carlson & Chemtob, 1990). This is also reflected in communicating in writing where there are different conventions for laying out formal letters compared to friendly emails or texts. In fact even the presence of another person will influence the display of multimodal expressions. Although some researchers such as Kraut and Johnson (1979) and Fridlund (1991) see use of expressions primarily as signals, even when there is no audience present to view them. Ekman (2003) explores this idea, suggesting that some expressions occur when people are alone, for example in response to environmental stimuli such as thunder or a beautiful a sunset. He argues that some expressions are more likely to occur during interaction because the primary function of emotion is to mobilise an organism to deal quickly with interpersonal encounters, in response to others, or the environment, or in order to express oneself.

11. Social and cultural perspectives vary from one cultural or social group to another. Minsky (2006) suggests that emotions are complex control networks that enable us to function effectively in the social world, for example by suspending our critical faculties for those we love. He gives the example of functions that normally generate disgust but are acceptable in loved ones, such as changing a baby's nappy. A major strand of Western philosophical thought viewed emotions as being opposed to reason and although those who study emotions no longer hold this view, it may still be found embedded in popular culture and use of language. For example in Western cultures, fictional characters are often characterised as able to make rational decisions *because* they are devoid of emotion, such as Dr Spock in Startrek (Gibberman, 1991). We have many examples of use of language in Western cultures that show this stance, such as 'keeping a cool head' that suggests an absence of emotion (or at least of its portrayal) is required for management of critical situations.

We need to consider many aspects that may vary between different societies and cultures such as transference of emotions between people and how imitation of others is important in understanding (Hurley & Chater, 2005). Emotions are induced by social contagion (Jones & Jones 1995) and collective climates, such as the aftermath of 9/11 when the World Trade buildings were destroyed in New York. Emotions help individuals to manage their responses to such tragedies (Fredrickson, Tugade, Waugh & Larkin, 2003). Emotional feelings may be diminished or absent in those who for physical or neurological reasons are unable to imitate (Goldman, 2005). Some emotions may be learned, for example aggression (Phillips 1983). Social support may buffer emotional experience (Wade & Kendlar, 2000) and mood memory can alter experience of emotion (Eich & Macaulay, 2006) and judgement can be altered by mood congruency (Isen, Shalker Clark & Karp 1978).

Culture is an important variable. We follow complex culturally dependent social rules to convey both emotional and other information and to serve social functions. Some emotional reactions are cross-cultural (Goeleven, De Raedt, Leyman & Verschuere, 2008), but many non-verbal cues alter their meaning in different cultures (Wierzbicka 1999). Rules of politeness vary; for example, eye contact denotes attentive behaviour in Western cultures, but not in oriental cultures where, in some interactions, it may be considered rude. Expectations based on a cultural and stereotypic basis may influence the way emotions are shown, or reported, or learnt. Matsumoto and Kupperbusch (2001) found that American (compared to Japanese) females showed less negative expression in response to an unpleasant film if an authority figure was present.

Development of affect and emotional responses is also bound up with culture, and particularly with sensitivity to the behaviour of the mother and to maternal attachment, with babies using the mother's emotional expressions for social referencing (De Wolff & Van Ijzendoorn, 1997).

Emotional values are culturally determined and displays of emotion are governed by different social rules and expectations in different cultures. For example there are big differences between collective cultures of the east and individualistic cultures of the West, or between in-groups such as youth cultures. Eid and Diener (2001) found many inter and intra-national differences. Kleinsmith, De Silva, and Bianchi-Berthouze (2006) found it necessary to build separate cultural models for affective posture recognition.

Early anthropological studies of emotion by pioneers such as Margaret Mead (1928) and Jean Briggs (1970) found differences in the isolated societies they studied, for example Mead found emotions in adolescence varied between Samoa and the Western world and Briggs found that in the Utku Inuit society, expressions of anger and aggression almost never occurred, despite Ekman describing these as universal 'basic' emotions, recognised by facial expressions. This may have been due to differences in methods and coding. Ekman's original work depended on asking people from different cultures to identify emotions from still images. Matsumoto & Ekman (2008) replicated Ekman's study of rating facial expressions and found cultural decoding rules varied. Recent studies on recognition of universal expressions rely on recognition using video rather than still images, enabling recognition of dynamic movements and micro expressions.

12. Appraisal Appraisal theorists offer compelling arguments that appraisal of emotion is an important variable. They suggest that the way people think about or judge their feelings actually alters them. Arnold (1960) called appraisals 'sense judgements' that are "direct, immediate, non-reflective, non-intellectual (and) automatic". Levenson describes how affect programs activate expressive motor and experiential systems – e.g. prosody in speech (Levenson**,** Ekman, & Friesen, 1990). Fridja (1986) describes action tendencies and appraisal patterns which link personality physiological state, learning history, context etc. Le Doux (1996) and Zajonc (1980) argue that emotions operate faster than appraisal possibly could and point out that it is possible to elicit emotion without the subject being aware of the stimuli. Lazarus (Lazarus, Speisman, Mordkoff & Davison, 1962, Lazarus, 1993) showed that if you change the appraisal pattern – then emotion will change, for example he observed responses to a gruesome film that were moderated by the type of information given to the subjects. Catanzarao and Mearns (1999) used people's ability to alter their appraisal in this way as a basis to develop a method for measuring psychotherapeutic outcomes. Scherer, Shorr & Johnstone (2001) reviewed the different appraisal theories and their sets of appraisal components or dimensions e.g. control, certainty or effort that different theorists describe.

While we cannot always be sure exactly how different people appraise different stimuli, there are some sets of stimuli that have been developed and responses validated so we can be confident that they reliably elicit the same emotion. For example Velten (1968) mood statements consist of a list of statements such as 'My favourite songs keep going through my mind' or 'Sometimes I've wished I could die' that are respectively correlated with positive and negative affect. There are inventories of pictures and sounds (Bradley & Lang, 2007; Verona, Patrick, Curtin, Bradley & Lang, 2004) that are validated to invoke emotional responses such as a feeling of disgust in response to a picture of a dirty toilet. Some aspects of systems have been found to elicit a feeling of frustration, such as having time pressure, interrupts or delayed system responses (Riseberg, Klein, Fernandez & Picard, 1998; Scheirer, Fernandez, Klein & Picard, 2002) or poorly designed interfaces (Ward, Cahill, Marsden & Johnson, 2002).

Implications: There are a number of individual and contextual variables that might impact on my work and that I will need to consider. People vary in their communicative skills and strategies and in the social rules they follow, dependent upon differences in their culture, social context, peer group norms, age, health, personality, sex and gender and individual make-up etc. There may be too many variables for me to control all of them in my study, but I should at least observe and measure as many as I reasonably can and consider how they might impact on any findings. The placing of emotion in a social context is part of my user centred perspective and as I conceive human communication as a primarily social behaviour, it makes sense to consider methods and lessons from the social sciences in my approach.

2.4. Terminology

2.4.1. Difficulties of Definition

It is important to establish a rationale for using terminology. If thought and emotion are inseparable then it could be said that emotion is everything we say and everything we do. As our knowledge about emotions changes, and depending on context, our use of language changes. The current standard dictionary definition of '*emotion*' is 'moving of the feelings, agitation of the mind, one of the three groups of the phenomena of the mind – feeling – distinguished from cognition and will' (Chambers Dictionary, 2004). Chambers definitions of other important terms associated with 'affect' demonstrate how difficult it is to clearly articulate what is meant by emotion:

- **Affect** - a feeling or emotion as distinguished from cognition, thought, or action;
- **Mood -** a longer lasting affective state;
- **Emotion** - an intense feeling; a complex and usually strong subjective response, as love or fear; a state of agitation or disturbance;
- **Feeling** - sensation perceived by the sense of touch; an indefinite state of mind; an affective state of consciousness, such as that resulting from emotions, sentiments, or desires; an emotional state or disposition; non-intellectual or subjective human response;
- **Subjective** - a feeling proceeding from or taking place within an individual's mind.

Until recently the concept of emotion has been poorly defined and seemed intractable. Mandler (1975) suggested it is pointless to search for the theory or the definition of emotion, as there are so many varying viewpoints and starting points. Fehr & Russell (1984) said ‘everyone knows what an emotion is, until they’re asked to give a definition, then it seems no-one knows’. Fridja (1986) stated that ‘a definition of emotion can only be a product of a theory; it can thus be reached only at the end of the investigation’. Kleinginna and Kleinginna (1981) suggested that no matter how far the study of emotion goes, it is impossible to find one correct and unique definition. They looked at 92 definition and nine skeptical statements about the concept of emotions and concluded that there is little consistency and too much vagueness. They suggest a more comprehensive definition:

> “Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can:
> (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure;
> (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes;
> (c) activate widespread physiological adjustments to the arousing conditions; and
> (d) lead to behavior that is often, but not always, expressive, goal directed, and adaptive." (Kleinginna and Kleinginna, 1981 p. 355).

Terminology used in emotion research has been influenced by the theoretical and physical basis and by the interdisciplinary nature of research. Ortony and Turner (1990) define emotions as qualitatively distinct phenomena including both feeling states and cognitive components including appraisals, intentions, and desires. Izard (1992) in his differential emotions theory describes emotional experience as

> “a feeling state or motivational condition, a direct and immediate product of the particular neural processes associated with that emotion” (p561).

He draws a distinction between motivation (emotional state) and motive (cognitively articulated goal), while Hinde (1985) felt ‘expression of emotions’ was misleading and that behaviours were more simply an expression of an internal state. Clore, Ortony and Fosse (1987) describe the psychological foundations underlying our affective lexicon.

The word ‘emotion’ has been used to describe both conscious and unconscious phenomena, and a spectrum of events from lifelong personality traits, sustained moods, feelings or states and fleeting momentary experiences. Mood is often seen in contrast to shorter, stimuli dependant emotions, as having a longer span and related to a more generalised stimuli, making it more difficult to find the source than for it than for a brief feeling of emotion. A fine line exists between stimuli that cause emotional feelings, emotional feelings themselves, resulting emotional expressions and phenomena arising from feelings and expressions. For example ‘pain’ is something that can be given or received. Also included in taxonomies are names for complex emotional conditions such as confusion, boredom etc. that might have several component parts and lie on a spectrum of meaning.

Researchers have tried to categorise emotions using descriptive words and taxonomies. For example the Linguistic Inquiry and Word Count website lists 2290 words and word stems with emotional classifications (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007). Baron-Cohen developed a taxonomy of 421 emotion words organised within 24 categories (Baron-Cohen, Golan, Wheelwright & Hill, 2004). People also use linguistic devices such as similes or metaphors to describe emotions, e.g. ‘as cool as a cucumber’ ‘pleased as punch’ ‘I just blew my top’ ‘going ballistic’ etc. In practice it is difficult for individuals to compare emotions using word labels because a word may have different meanings in different dialects, or for different people, at different times, for example one person's expression or observation of 'miffed' might equate to another person's idea of 'upset, angry or furious' (Brown & Silzer, 2001). People may be describing blended or mixed emotions or different degrees of emotion. People often use synonyms and metaphors to describe emotions. The task is complicated because we have so many words for emotions – estimates vary from two to six thousand in English. Some words translate into different languages but others do not translate exactly, or carry the same connotations and subtleties (Wierzbicka 1999, Brown & Silzer, 2001). Some meanings are culturally dependant, for example 'love' is rated positively in English – but there is a negative category of 'sad love' in Chinese. The same word may have quite different meanings when used in different domains or contexts. Etymology means that meanings of words change over time, for example 'worried' can mean to be savaged by an animal, but that meaning is no longer commonly used. In the technological domain neologisms (new words) and novel uses of words in English are frequent events, as demonstrated by Wordspy (McFedries, 2009).

Implications: Definition of emotions and related states is a complex area. Ambiguous definitions can lead to misunderstandings. Clarification of use of terminology is important.

2.4.2. Definitions Associated with Affective Computing

Rosalind Picard (1997), in her book 'Affective Computing', used the term 'affect' to encompass the entire spectrum of emotions, which includes feelings, moods, sentiments etc. She uses a nice analogy with Mount Everest, suggesting that arguments about exact definitions need not prevent research progressing. While it may be difficult to define if any particular rock constitutes a part of Mount Everest or not, it is still feasible to discuss facts about the mountain such as its height, climbers' achievements etc. In the same way we all have a broad understanding of the concept of emotion and we can discuss issues related to it.

One important distinction raised by Picard (1997) is the distinction between inner emotional experiences and the outward emotional expressions that people may use to convey messages about their emotional states. She uses the term 'emotional expression' to describe what is shown via bodily systems, "what is revealed to others either voluntarily, such as by a deliberate smile, or involuntarily, such as by a nervous twitch' (p25).

The use of 'affect' is criticised by Ortony who suggests that in its pure meaning, 'affect' relates to a measurable valenced phenomena.

Theorists have proposed useful conceptual definitions based on their work for example Scherer's was found useful, though incomplete by the Humaine[1] review (Cowie, 2007):

> "episodes of massive, synchronised recruitment of mental and somatic resources allowing to adapt or cope with a stimulus event subjectively appraised as being highly pertinent to the needs, goals and values of the individuals" (Scherer, Wranik, Tran & Scherer, 2004, p 10).

'Design features'	Emergent emotion	Interpersonal stances	Moods	Attitudes	Affective disposition
Impact on behaviour	+ +	+		+	+
Intensity	+ +	+	+	+	
Rapidity of change	+ +	+ +	+		
Brevity	+ +	+	+		
Event focus	+ +	+			
Appraisal elicitation	+ +				
Synchronisation	+ +				

Table 3. Genera of emotional states (after Scherer 2004) from Humaine deliverable document D3i Emotional life: Terminological and conceptual clarifications

With emotion theorists and researchers struggling to supply and agree on a common definition, the Humaine collaboration put significant effort and debate into the issue resulting in a deliverable that attempts to clarify the current state of language and move towards a consensus for use of language and terminology in this area. Table 3

[1] Humaine (Humaine website, 2009) is an EU-funded network of excellence that made a co-ordinated effort to come to a shared understanding of the issues involved in emotion research, and to propose exemplary research methods in the various areas.

shows how they suggest different design features might relate to various concepts. The Humaine deliverable on terminological and conceptual clarifications (Cowie, 2007) describes emotion in a 'strong' everyday sense as:

> "a state where the person's whole system is caught up in the way they react to a particular person or situation – which may be in reality or in their mind. Specifically, it involves distinctive positive or negative feelings about the people or situations involved; it involves impulses to act or express yourself in particular ways and avoid others; it involves distinctive changes in your body, for instance in your heart rate or tendency to sweat; it doesn't usually last very long – it comes on quite quickly, and dies down reasonably soon (unless there is something very unusual happening)**."**

Humaine contrasts this definition with a more specific working one with three strands:

- "*Emergent emotion* consists of episodes that fit Scherer's definition and the corresponding experiential description;
- *Pervasive emotion* is what runs through (or colours) the rest of emotional life;
- *Emotionless* states involve neither emergent nor pervasive emotion."

Implications: Affect is shown by any expression (made up of multimodal expressive parts) and modified by functional interactive context. I need to decide on and use unambiguous terminology to describe complex concepts relating to emotion. For clarity of concept and functional use in my study, (carried out before the Humaine deliverable that sought to clarify this area) I adopt Picard's use of the word 'affect', to encompass the spectrum of emotions and 'affective computing' to cover the spectrum of emotional systems. I differentiate inner 'emotional feelings' and outer expression of emotion through emotion related behaviours. I use the term 'multimodal expression' to describe the broader range of communicative behaviours that may or may not be primarily concerned with communication of emotions. In my discussion and conclusions chapters (Chapter Seven and Eight) I will return to the issue of complexity of affective expressions and their description.

2.5. Affective Computing Applications

There are lots of ways emotion could be used in computing that might be fun or useful…..or threatening. It is likely that different realisations of the concept will develop in different ways in different platforms for different domains and purposes.

2.5.1. Automatic Emotional Detection Systems

Systems to detect the emotional states of users are in their infancy, but with considerable investment and interest in them. Systems are being developed to detect various aspects of emotion in various ways. As systems for image recognition, biosensors and pattern recognition are becoming wireless, more powerful and more pervasive, the potential for automatic detection of emotional states is progressing rapidly, for example see review by Peter & Herbon (2006); Humaine report (Gelin, 2005); survey by Zeng, Pantic, Roisman & Huang (2007).

Automatic face recognition and facial expression recognition is under development (see review by Fasel & Luettin, 2003) for example there is an automatic version of Ekman's facial action coding system (FACS). The Facial Expression Analysis Tool (FEAT) uses pattern recognition, fuzzy rules and learning algorithms to automatically measure facial expressions. It claims results comparable to a human expert. El Kaliouby and Robinson (2004) take this a stage further with their 'face reader' that tracks facial expression in real time and gives a graph to show probability of up to six complex emotional states, learnt from Baron-Cohen's (2003) Mind Reading taxonomy. Other systems detect gaze direction, and pupil dilation. An alternative to visual detection systems involves the use of heat maps or sensors to detect muscle activation, for example to muscle movement during blinking (which might be indicative of lying or fatigue). An example of this is Lal and Craig's (2002) use of electroencephalography to detect anxiety and fatigue in driving simulation experiments.

Speech analysis systems consider the choice of words and appraisal of the meaning of words. Vocal characteristics of speech such as emphasis on parts of words and intonation are also important (see review by Scherer 2003). Analysis of voice has been used to detect vocal strain for professions such as teachers and call centre employees (e.g. Lehto, Laaksonen, Vilkman & Alku, 2008) and might also be used for analysis of the emotional state of incoming callers. One controversial example is a system to detect when callers making benefits claims are truthful or telling lies (TUC, 2007). Fear is another characteristic that can be detected and could be used to improve public safety issues (Clavel, Vasilescu, Devillers & Richard, 2008).

Postural and gesture detection are proving useful additions in the field. For example, Kleinsmith and Bianchi-Berthouze (2007) developed automatic recognition models to discriminate between levels of affective dimensions based on low level postural features of static postures. Berthouze is exploring models of posture and gait to help doctors to detect pain in their patients. Kiefer has developed a system that uses a camera to recognise hand motion (based on visual recognition of skin tone) that can be used to control emotional music mixing. One potential use for this tool would be for people with limited movement and mobility, as a musical instrument (Kiefer, Collins & Fitzpatrick, 2009).

Positioning in time and place is another feature that can be detected and add emotional meaning to human behaviour. For example intelligent surveillance systems can give alerts if someone is acting in a furtive manner, for example walking or climbing over a wall in a prohibited area or at a prohibited time. Systems that log various biometrics can help realise this (e.g. Bullington, 2005; Wang & Fang, 2008).

Smell and taste are important in emotions – but often operate at a sub conscious level, for example in the detection of the smell of pheromones or sweat, but detecting these does have potential in affective computing and Lindeman and Noma (2007) have developed a multisensory classification system that could be used to develop augmented reality systems. Suslick and Rakow (2001) have been developing an artificial nose that can recognise certain gases. Lavigne, Savoy, Clevenger, Richie, McDoniel, Yoo, Anslyn, McDevitt, Shear, and Neikirk (1998) began development of an artificial tongue. These technologies are already being used to give initial responses e.g. for testing foodstuffs and detecting substances such as bombs and drugs and these might add to multisensory detection of emotions in the future.

Touch is also important for emotion, for example we can tell a lot about a person's emotional state from the texture and strength of a handshake or a hug. Dana Kirsch (1997) developed a sentic mouse. This was an ordinary computer mouse, augmented with a pressure sensor to collect sentic data. When compared with other measures, including galvanic skin response, heart rate and self-assessment it was found that the sentic mouse could capture reliable emotional valence information. Commercial companies are using patterns of touch from click-streams during interactive TV and web use to profile emotional states of users (e.g. see BusinessWire 2002).

Brain activity is adding to our knowledge about emotional states and our ability to detect and differentiate various states with numerous studies using electro-encephalogram (EEG) and functional magnetic resonance imaging (fMRI) and most recently magnetoencephalography (MEG) scans to detect rapid emotion related activity. These are currently of limited use outside of experimental laboratories because they require subjects to be connected by wires, use very expensive technologies, and/or require subjects to keep very still during imaging procedures. However sensors are under development that can function wirelessly and remotely (e.g. Prance, Beardsmore-Rust, Watson, Harland & Prance, 2008). Meanwhile there are systems that claim to detect 'recognition' waves in the brain that could be used to screen for terrorists boarding planes (SSRM Tek, 2009) and systems to detect depression via electrodes that pick up electrical impulses in particular areas of the brain associated with depressive feelings. These have potential to be used with new procedures such as deep brain stimulation to control mood (e.g. Mayberg, Lozano, Voon, McNeely, Seminowicz, Hamani, Schwalb, & Kennedy, 2005).

Complete contextual and temporal awareness by machines is not yet possible although it is essential for fully human-like capability. Humans will use context and time to help make judgements about a person's emotional state. For example, it might be quite unremarkable to hear a swear word in one setting but seem totally wrong or highly emotive in another. It might be normal for one person to use a swear word and not signify an emotional state for that individual, but for another person it might be very rare for them to use a swear word and be a sign of great stress. Someone might be running because they are trying to escape from a threatening situation or they might be running in a marathon. In order to accurately appraise an emotional state, a system would need to recognise an individual person and their current geographical and social context and then invoke and compare contextual and personalised knowledge about the individual and about past situations and settings and use inference to judge the probability of any one meaning.

It is likely that emotion detection systems will become both more powerful and more reliable as methods are developed and elements are combined to become multi-modal. For example Bullington (2005) proposes combining face recognition from surveillance camera networks with data from biometric sensors for security surveillance. Others have considered synthesis of various parameters, for example, prosody of speech (Schroeder, 2004) and facial muscle movements with heart rate, and electrodermal activity (Mahlke and Minge, 2008).

2.5.2. Types of Applications

The application of affective computing systems is a rapidly expanding field (for example the Companions Project[2] has published a set of papers to summarise the state of the art and future predictions e.g. of Embodied Conversational Agents (ECAs).

Virtual intelligent characters have been the focus of much work and are increasingly lifelike incorporating not only realistic facial expressions but other communication modes such as gestures and idle body movements. These ECAs are used in many applications from gambling games to health and exercise management applications (Wallis, Moore, Fagerberg, Cavazza & Wilks, 2006; Kenny, Parsons, Gratch & Rizzo, 2008; Halko & Kientz, 2009). The Semaine project (2009) aims to build a sensitive artificial listener.

Robotics There are a number of research groups building robots that display facial expressions (Canamero & Fredslund, 2001) or artificial personalities, developed

[2] COMPANIONS (Companions website, 2009) aims to change the way we think about the relationships of people to computers and the Internet by developing a virtual conversational 'Companion'.

using machine-learning techniques. Steve Grand who invented ‘Creatures’ is working on ‘Lucy’ a baby orangutang robot he hopes will get to nursery school stage (Grand, 2004). There are markets for robotic pets such as Tamagochis or the Sony Aibo. There is potential for robotic companions for the elderly or isolated individuals, although the social acceptability of these is not established and may vary with culture. Researchers are tackling issues around the social acceptability of these robots (e.g.Dautenhahn 1994; Breazeal, 2002). Considerable investment in creating social robots (see Intini, 2005) is ongoing, using emotion models to enhance machine intelligence and aiming to develop systems that can communicate with humans effectively and give an illusion of life. (e.g.Ushida, Hirayama, and Nakajima,1998; Satake, Kanda, Glas, Imai, Ishiguro & Hagita, 2009).

Marketing applications and web design now often use avatars and speech interfaces or human images to help promote services and sell goods and to inspire trust in users (e.g. Steinbrück, Schaumburg, Duda, & Krüger, 2002). Affective systems could detect and respond to levels of interest and engagement, matched to metrics of customer history of purchases and interaction, resulting in customised web pages or messages and enhanced experiences such as adapting choice of content or timing of delivery to help users to reach relevant or interesting information more quickly. Usability and satisfaction with computer applications might be enhanced, but user control could lessen. Lie detection using polygraphs is unreliable so affective systems might prove useful for interrogation. In call centres voice analysis is now sometimes used to help manage calls, e.g. to detect probability of callers being angry or lying (TUC, 2007). This is a controversial area as it has many implications for privacy and ethical behaviour. It seems that using feelings to aid decision-making can have a positive effect on individual satisfaction with decisions made. Schwartz (2004) carried out an interesting study tracking the emotional state of groups of participants who were set problem solving tasks that such as a car buying task that involved either digesting a large amount of information or relying on a much reduced amount of information and so relying more on ‘gut’ feelings. Those who relied on ‘gut’ feelings proved much happier with the decisions they reached.

In gaming and multi player games development there is huge investment with novel interfaces. For example, recently developed interactions include some that use blowing on the screen. Sony incorporates some facial and gesture recognition in their Playstation eye-toy technology. This could be used to enable avatars to mimic the facial expressions and non-verbal behaviours of their players, or to detect excitement or boredom and adjust levels and offer help to the player accordingly (Broersma 2003). An opponent's avatar could also use emotional expression. Games could be designed so that some desirable emotions are rewarded. For instance if you demonstrate certain emotional factors like good concentration, on feeling brave and showing no fear, then you could get extra points or game advantages. Affective applications may improve playability, usability and likeability of products.

Educational settings Online education is increasingly popular such as open courseware, which could use the links between affect and cognition to improve learning opportunities. These could be delivered through interactive TV channels (Dillon & Jobst, 2005). Trust and engagement levels could be assessed and the system adapted to improve effectiveness and to tailor methods to individual learning styles, for example simplifying material or repeating items poorly attended. As well as in virtual education settings, the balance of negative and positive affect is of interest in the real classroom. Alsmeyer, Luckin & Good (2008) used 'subtle stones' for pupils to feedback their emotional states to their teacher. There is evidence that both positive and negative emotional states can support learning, for example regret can drive motivation to learn (Saffrey, Summerville & Roese, 2008). The delicate balance and optimum requirement for switching between various levels of positive and negative states is not fully understood as yet.

Smart homes and office environments are under development responding to emotional states or multimodal expressions to select levels of lighting, heating, background music, etc. in order to enhance or create a desired mood. In retail settings sensors could detect emotional signals such as postural cues or even analysis of shopping baskets (that might contain 'comfort' foods for example). Targeted advertising could then be directed at individuals depending on their emotional state and intended to persuade them to buy products, depending on their mood.

A number of therapeutic applications have been developed using affective design to help overcome phobias such as fear of flying, fear of heights or spiders, or public speaking and post traumatic stress disorder. One application helps teach children how to deal with bullying by using empathic cartoon characters to teach social communication skills (Hall, Woods, Hall & Wolke, 2007).

Monitoring emotional states for health or other management purposes is also a development area. Sensors, possibly embedded in clothing could detect physical signs and alert users. For example, warning of over exertion, low blood sugar levels or the onset of agitation for those prone to high blood pressure. The onset of episodes of rage could be detected to support anger management courses. Virtual agents could offer counselling or health advice (e.g. Bickmore, Schulman & Sidner, 2009; Relational Agents website, 2009) and Automatic detection of affect could be used as a teaching aid for those with attentional deficits. There is already a system for cars to track blink rates that relies on detecting frequency of blinking, as you tend to blink more if you are getting sleepy. It is hoped it will prevent accidents due to drivers falling asleep (Fletcher & Zelinsky, 2007), although it cannot yet deal with real world issues such as people wearing sunglasses when driving. Trials are under way using facial recognition in conjunction with CCTV footage to try and detect individuals likely to commit crimes (e.g. Shan, Chen, Sanderson & Lovell, 2007).

Devices to teach, enhance or manage emotional and social skills and thus enhance emotional relationships are also potential development areas. There is a move currently towards applications that offer some emotionally related information, and allow users to interpret that additional information in their own way and for their own purposes. Earrings could whisper a reminder of someone's name, recognised from their voice profile, or special glasses or contact lenses could identify and name faces and remind you of their emotional state on prior occasions (Picard, 1997). Systems could assist in emotional labour, for example Boehner's use of an interactive version of a Miro painting (Sengers, Kaye, Boehner, Fairbank, Gay, Medynskiy & Wyche, 2004) that helps reflection on the emotional state of co-workers or the affective diary (Lindström, Ståhl, Höök, Sundström, Laaksolahti, Combetto, Taylor, & Bresin, 2006). Fleck and Fitzpatrick (2003) used a Senscam (a body worn, continual use camera) for teachers to create a reflexive diary of their teaching practice. Teaching emotion skills to those who find them challenging, such as to individuals on the autistic spectrum is another area of interest. For example 'The Transporters' (Baron-Cohen 2009) is a series of videos featuring emotional cartoon train characters for use with children on the autistic spectrum and Picard (2009) has been working with autistic individuals who are helped to reflect on interactional issues by feedback from a wrist worn galvanic skin measurement device. You could indicate your mood by selecting an 'affecticon' to display to another person, which might be an image, background colour or icon in keeping with your mood. Emails or text messages can have background colours that indicate your mood to the receiver, or small images of particular multimodal expressions could be portrayed by photos, videos or avatars e.g. eMoto (Sundström, Ståhl, & Höök, 2005). On-line presence can be indicated by icons such as in 'Babble' (Erikson & Laff, 2001) and could be amended to show mood by degrees of colour or animation.

Emotional toys such as Tamagochis (Tamagochi, 2009) or a dinosaur 'Pleo' (Pleoworld, 2009) that reacts to stroking and voice are already commercially available based on earlier developments e.g The Affective Tigger (Kirsch, 1999). Toys have been found very motivating to children, encouraging them to attend to them conscientiously at least in the short term, but there are concerns they may teach children only shallow emotions as they suffer no real harm when neglected e.g. Ribi, Yokoyama, & Turner's study of children interacting with robotic Aibo dogs (2008).

Implications: It is important to consider the general perspective within which the development of affective systems takes place and to consider the likelihood and importance of their development within our society and culture in general.
For my study I need to decide on a suitable application that will reflect the kind of systems already under development, or envisioned for the future. This is challenging, as systems are in their infancy. A game environment of some kind might provide a good setting.

2.5.3. Limits of Computer Recognition and Display of Emotion

In affective computing there has been a strong bias for overcoming technical challenges, concentrated on limited communication channels, for example, emphasis on complex recognition algorithms. Responsive systems may involve use of models, artificial intelligence, neural networks and machine learning techniques, so that the system develops a degree of emotional intelligence. Many prototype systems to date have concentrated on either recognition or display, rather than incorporating both.

Currently there is a growing interest in how affective recognition and displays can be incorporated into computer applications set in social contexts, to either induce or display the user's empathy or emotion, or both. This is really cutting edge technology and researchers in this area readily acknowledge that systems may disenchant users if they claim to accurately display or read emotions and then fail to perform well. One approach is to build systems to enhance, rather than replace human skills and to rely on users for interpretation (e.g. Sengers, Boehner, Mateas, & Gay, 2008).

Despite an explosion of research and development in affective computing, capabilities do not yet approach human abilities to detect and respond to multimodal expressions. Considering the theories and models that require an appreciation of levels of processing and context, to even approach human capability they must:

1. Perceive and recognise affective displays of multimodal expressions. The state of the art systems at present at best only combine recognition of a limited number of modes, for example el Kaliouby & Robinson's application (2004) detects head gestures and facial expressions.
2. Understand the context that the displays are made in and the social rules that apply and synthesise this. Current systems make no attempt to process this type of information.
3. Cognitively assess emotional displays, to reach an understanding of meaning. While some systems claim to interpret emotional expressions, these claims are largely spurious. However an interesting trend in systems is to try to provide added information to humans in order to assist cognition and understanding of meaning. For example, el Kaliouby & Robinson's (2004) application gives

probabilities of various emotions to assist cognitive interpretation by the system user.

4. Manage any feelings generated by that understanding. Current systems to not actually manage feelings, although feedback from systems might assist an individual to manage their own feelings e.g. see Picard (2009).
5. Select responsive modes and respond appropriately, with multimodal expressions. ECAs and robotic companions that try to show appropriate emotional responses are under development, and offer exciting promise, but these systems are in their infancy. They are beginning to include more and more types of non verbal communication (e.g. deMelo & Gratch, 2009) or see the 'Greta' agent (de Rosis, Pelachaud, Poggi, Carofiglio, & De Carolis, 2003).
6. Monitor own expressions and responses of listener and modification of use of emotional expressions accordingly. Current systems do not yet do this.

There has been little research into what users prefer in different domains, applications or platforms. Displays could involve:

- using behaviour of artistically designed cartoon characters, such as Disney cartoons;
- using verbal (spoken or written) emotive messages;
- using images or avatars, giving verbal or non verbal communication messages;
- using adaptive pace or content e.g. for web browsing or learning applications;
- using a shared code of behaviour – e.g. a colour or image or behaviour that has agreed meaning;
- reflection, mimicking or enhancement of user's multimodal expressions e.g. in game environment or cooperative work system;
- a system with a whole artificial intelligence personality e.g. an artificial companion.

There are emerging issues about what affective systems people find acceptable and the problems that arise with avatars that are very realistic and yet leave the user feeling uncomfortable. For example Masahiro Mori describes how people have negative reactions to facsimiles that try to appear very lifelike. He calls this dip in positivity the “uncanny valley” (Mori, 1970, 2005). This has presented issues for developers of ECAs seeking to implement ‘believable agents’ (e.g. Knoppel, Tigelaar, Bos, Alofs, & Ruttkay, 2008).

Implications: Examination of extant prototype systems informs us that any but the most simplistic efforts at recognition and display of emotional expressions is beyond the capabilities of any current application. As I seek to understand complex multi-modal communication and as even an inadequate attempt to build such a system would be very costly, a functional system is out of the question for the purposes of my work.

2.6. Human-Computer Interaction and Emotions

2.6.1. Attitudes to New Media

People’s feelings and attitudes to technology will affect their interaction. For example people often anthropomorphise computers and applications (Marakas, Johnson & Palmer 2000). There are many examples, anecdotes and urban myths about episodes of frustration or rage triggered during human-computer interaction, from threats to smashing terminals, or attempting to barbecue an electronic mouse (Kent, 2009; Tickner, 2004).

In an interesting series of experiments Reeves and Nass have shown that people tend to treat new media as if they are real people or things. For example they found that participants asked to try out and then give their views about a computer application were kinder in their views when responding on the same computer, than when giving their views given on a different machine or by using pencil and paper (Reeves & Nass, 1996). Kaiser & Wherle (1996) found that people do show emotional facial expressions and report emotional feelings during online game playing. Only a tiny amount of human attribute can give presence and artists or designers may supply this intuitively. Research comparing multimodal expressions of individuals interacting with real and virtual characters is in its infancy.

Humans intuitively adapt and develop new rule systems for new situations. For example, users have developed new methods, terminologies and social rules to communicate affect in social networking using technologies. Wallace (1999) describes the use of new conventions for spellings, formatting, layout and emoticons in emailing and messaging. Ever more novel rule systems for emerging affective technologies are likely. Studies are beginning to assess how users will adapt their emotional behaviour during interaction. For example, Paiva, Andersson, Höök, Mourão, Costa, and Martinho, (2002) found people intuitively manipulated Sentoy, a doll like figurine, to reflect their emotional expressions, which in turn controlled an interface. Isbister, Höök, Sharp, and Laaksolahti (2007) went on to develop a novel interface – the Sensory Evaluation Tool - for people to communicate their emotional feelings, by manipulating textured objects.

If the emotions that people feel are dependent on what is physiologically happening to them, and if their physical behaviour influences the way they experience emotion, this has implications for design of human computer interaction systems. It leads us to question exactly what physical movement and / or expression of emotion is invited by a system and how that might influence individual experience of using a system and how users feel while interacting with it.

Implications: As systems that can recognise and / or respond to affect are in their infancy, we are only just beginning to gather evidence about how people actually interact with affective systems and their attitudes to them. We need more research to identify and understand the underlying science.

2.6.2. Ethical Considerations

Data about user's emotional states could be easily collected by future technological systems and misused for psychological manipulation. Futurists predict that the development of such technologies is inevitable and that we should take collective responsibility and consider guidelines for their ethical development and deployment. Ray Kurzweil (Hoover Institute, 2009) suggests:

> 'I think the answer is a set of ethical standards of responsible practitioners….. I think it's a collaboration between the technologists and society. I mean, they say war is too important to leave to the generals. I think technology is too important potentially to the technologists. I think it's got to be a whole social discussion.'

There has been little research into surveying attitudes to affective computing. In a small pilot study run in conjunction with a museum display about affective computing, I found that participants found it hard to grasp principles about affective machines, and that questions about affect needed very careful design to be comprehensible to the general public. People asked what I meant by 'emotion' and wanted specific emotions to be defined. They seemed better able to understand 'feeling' as a general term.

With systems still in their developmental stages, it is difficult for ordinary users to understand the concepts, and questions in order to give an informed view, especially if researchers themselves do not use associated terminology consistently. Schuller Müller, Hörnler, Höthker, Konosu and Rigoll (2007) developed a prototype system that could recognise some emotions, via some combined speech and gesture recognition, with near human accuracy, and then asked users about their attitudes to the system. They found that only 8.3% thought it a 'must-have' feature, while 16.6% of their users felt that emotion recognition was a 'good thing' and a further 16.6% felt it was good, but raised fears and insecurities over issues such as privacy and security. Another third described feeling the whole issue was interesting, and 8.3% found the whole concept frightening.

Data might be collected without user's fully informed consent or understanding as to its use. Details about use may be hidden in the 'small print' of contracts, or users may be assumed to 'opt in' to uses. There are already concerns that some users provide content about their emotional states, along with other identifying data, to social networking sites, intended for sharing with a close circle of friends, but which can be easily accessed by others, for whom such information is not intended (such as future employers). In healthcare, electronic records containing information about emotional states is often submitted, in anonymised form, to large databases for secondary researchers to use. Another concern is the potential for merging data on emotional states with other sources of information such as personal details and purchasing or internet searching histories. There are concerns of privacy risks and criticisms of various companies who collect data on individuals, for example via placing ‘cookies’ onto PCs to collect ‘clickstreams’. For example, one company reportedly claims to provide ‘digital silhouettes’ of users by analysis of their click patterns during computer or TV use that can identify many pieces of personal information, such as gender, whether alone or in company, and some emotional states and they can use this information to target personalised advertising to users of interactive TV (Freed, 2001; Business Wire, 2002).

Persuasion is an ethically charged area, as persuasion might be used for nefarious as well as altruistic reasons. It seems it is only too easy to influence a user’s affective state. There is a long history of this in advertising (e.g. Mizerski & White, 1986; Percy, 2001; changingminds.org, 2010) for example with frequent use of dramatic emotional adverts on TV, (Moore & Harris, 1996). Persuasive technologies, such as those described by Fogg (2002) might be enhanced, as affective systems become more refined and ubiquitous.

Early efforts to build affective applications have produced interfaces that automatically detect and adapt to the user’s affect, and are designed to reduce negative affect and induce more positive affect (e.g. Klein, Moon & Picard, 2002; Picard & Klein, 2002; Partala & Surraka, 2004). Picard (1997) suggests there are four reasons for giving machines emotions:

- to emulate humans;
- to produce more intelligent machines;
- to understand human emotions better through computer modelling techniques;
- to produce applications which are less frustrating during interaction.

Tractinsky (2004) questions the wisdom of this suggest sthat systems should be configurable by the user. Well-designed applications should offer users choices. Lindgaard (2004) is concerned that systems aim to remove negative emotions such as frustration, which might play an important role in interaction. Negative emotions are a normal part of everyday experience and might play an important and useful role in some conditions, for example there is some evidence that negative affect can be a stimulus in learning contexts (Moridis & Economides, 2008), or be a necessary trigger for intentional change (Howard, 2006). Games and narratives in stories and films rely on emotional arcs that involve journeys through both positive and negative experiences to make them engaging (e.g. Freeman, 2003). Despite improvements in technologies such as electronic document management and speech systems and the availability of reasonably priced applications, paperless offices and speech interfaces have largely failed to catch on. Research is needed to establish how willing people are to share emotional communication with a technological system and they should have choice about whether to use affective systems, and in what conditions.

Implications: Ethical concerns are a real issue in affective computing and how users view such issues is not well established.

2.7. Summary

In this chapter I reviewed the perspectives of the background literature concerning theories of emotion, studies of the brain and models of emotion; contributions from human communication science in understanding both multimodal and individual variables; issues around definitions and use of terminology; surveyed affective systems and their limitations; and considered ethical issues associated with affective systems.

2.7.1. Summary of Implications

The literature shows that emotions are complex and operate on biological/visceral, neurological/ behavioural and psychological/cognitive processing levels. Brain imaging studies are gradually helping us to understand the complexity and exact mechanisms that underlie affect. Emotions are now recognised as being built through interaction and operating within a social context. The approach I take will reflect this understanding.

Human communication science contributes greatly to understanding the communication of affect. Early studies in affective computing neglected attention to the spoken word in favour of facial expression analysis (influenced by Ekman) but there is now an explosion of work looking at other modal aspects. Non-verbal expressions serve many other communicative purposes, in addition to sharing of information about emotional states. Uses and meanings of non-verbal signals are heavily influenced by the social context in which they take place. We integrate our perception of both non-verbal and verbal aspects to form our understanding of communication (Nygard, 2005).

As computing that relates to emotion is a relatively new area, the use of language is still emergent. Picard has been most influential in the field and I (along with many peers) adopt her use of 'affective computing' and I strive to clarify when I am referring to inner feelings or outward expressions of emotion. I refer to 'modes' of communication to describe the various non verbal or verbal means that are the toolbox of communicating messages, that might or might not include information about inner emotional states, or be intended as outward expressions of emotion. We need more research to identify and understand the underlying science of human computer interaction with affective systems and ethical concerns and how users view concerns, is not well established. Based on the evidence from the literature searches, a number of issues for this study have emerged:

Issue One: As shown throughout the literature searches and particularly in the 'Background' section (2.2), emotion is a hugely complex system. Use of terminology (section 2.4) is still under development within the field of research and is often not used consistently. Throughout this thesis, I will use the term 'affective computing' to cover the broad spectrum, but will try to clarify when I am referring to different aspects of emotion related behaviours or inner feelings.

Issue Two: As shown in the section on 'Human Communication' (2.3), emotion is expressed in complex ways, and via a combination of different modes. There is a lack of studies that consider multiple modes and complex naturalistic recognition of emotional expressions. In this study, I need to observe how users actually behave with affective systems, how, when and why they use emotional expressions and whether this varies under different conditions. I need to consider a range of multi-modal verbal and non-verbal signals. I need to focus on the outward expression of emotion and how it is understood and appraised. In chapter three I go on to describe methods for tackling this. Section 3.4 explains my methods for observing behaviours. Section 3.5 describes how a combination of qualitative and quantitative methods supports my observational work and section 3.6 describes the tools I use.

Issue Three: As shown in section 2.5, there are limitations to current state of the art systems, and consequently there is a lack of studies that use realistic recognition systems. These simply do not exist yet. In the following methodology chapter, section 3.3 explains how in my research, using a Wizard of Oz method enables me to simulate a realistic system.

Issue Four: Because fully working systems do not yet exist, there is a lack of empirical studies focusing on how users behave with systems and considering the user perspective. Early system development was led by what was technically most achievable rather than by what best suited users. As shown in section 2.6, because we do not yet have working systems, we do not yet know how users will act or feel with such systems or in what ways they might benefit or be impeded by forms of emotion recognition systems or types of emotionally responsive systems. To address this, in this study I need to:

- explore users' behaviour with, and views on affective systems;
- consider the context in which emotion related behaviours occur or vary;
- understand if using affective systems actually offers any help or improvement on standard systems.

In the following methodology chapter, section 3.2 explains how using experimental design will enable a focus on the user perspective and to compare conditions.

2.7.2. Overall Aim and Specific Objectives

My overall aim is to design a study to explore emotion in computing and to investigate this from the user perspective. It is apparent from the literature that there are clear issues to consider about developing computer systems that involve emotion. There is a real danger of costly, technology led development of applications that fail, because they do not meet user expectations or desires. I draw together ideas and findings from the literature search in order to devise an empirical study that takes a user-centred approach. The study will establish any user preferences and gains, as well as any issues for further consideration, which may be relevant to development of systems that use emotions in computing. My research aims to evaluate the potential for emotion recognition technology to improve the quality of human-computer interaction.

Resulting from the literature search and overall aim of my research, I must keep in mind the complexity of human emotional communication (issue one) and consider the multiple communication modes that people use to communicate their feelings (issue 2). As no satisfactory systems yet exist, (issue 3) I will have to simulate a fully working system in order to focus on the user perspective (issue 4). By varying the conditions of use of an affective system for different groups of participants, I will be able to address the bullet points described in issue four, relating to the user perspective, in order to achieve the objectives arising, which are to:

1. Establish the extent to which people will naturally express emotions when they know they are interacting with an emotion-detecting computer.
2. Identify the conditions under which the application of emotion detection can lead to improvements in subjective and/or objective measures of system usability.
3. Establish if and to what extent participants are willing to adapt their behaviours during interaction with an apparently affective system.

In Chapter Three I explain how I selected methods to address the issues raised here and in Chapter Four I translate these objectives into a set of hypotheses and describe the experimental procedure used to test them.

3. Methodology

3.1. Introduction

This chapter explains how the review of the literature and the aims and specific objectives that resulted from it led me to selection of an experimental design and Wizard of Oz method for my study. It explains the choices of methods and the decisions made in order to efficiently observe and identify and measure behaviours relevant to expression of emotion, and the tools selected for use.

3.2. Experimental Design

3.2.1. Rationale

There are various possible methods of researching the issues identified for consideration. A series of longitudinal case studies of a real world application with naturalistic observation would be ideal for studying details and complexity of ongoing interaction with an affective system. But in the absence of a rigorous, working, affective computing system this would be difficult or impossible to design. A survey could yield information from a larger number of people but it is impossible to effectively question people about use of affective systems when they lack experience of using such systems and in any case widespread use of terminology and language is not developed enough to easily discuss emotional systems. Interviews would give more in depth information, but are similarly impossible when there is no real world experience of affective systems. Ideally if I opt for an experimental design I would conduct the experiment in a real world setting, in the field, but as realistic systems do not yet exist, that is not possible and I have to use a laboratory setting. Quasi experiments using naturally occurring situations, such as comparing people in the real world, who do or do not use affective systems is similarly not feasible.

3.2.2. Method

Experimental design is a method of enquiry based on gathering observable, empirical and measurable evidence that is subjected to specific principles of reasoning (Ryan, 2007). This suits my needs as I have identified a lack of empirical evidence in the literature and want to focus on an observable and measurable behaviour - how users show emotional expressions when using an affective system. Experimental design allows comparison of groups under different conditions and selection of methods to measure their behaviour. Enough participants can be included in each group (decided by a power calculation) to ensure rigour and statistical power for analysis and any findings can be subject to qualitative and / or quantitative statistical analysis.

3.2.3. Disadvantages

Disadvantages of this method are the lack of a real world setting. From a laboratory study, I can only infer how any results may pertain in the real world. People might behave quite differently in the artificial environment of a lab and behaviour might be distorted, so findings might not be ecologically valid and so less useful. For example participants might have emotional feelings that they express related to the lab setting, so that what I detect and measure is due to the setting rather than to the dependent variables. Critics of experimental design maintain that observations and findings are shallow because they are not the same as the rich and varied variations of the human condition encountered in the real world. Controlling the experiment may result in also controlling the behaviours to be measured. Controlling various aspects of the experiment takes time and resources, so I can only include a certain number of participants (unlike surveys or questionnaires that could be sent to a wider group).

Focus on control of the variables identified might result in minds shut to other issues, so that some variables are included that we are unaware of and that we are not controlling. For example, in the case of my study, I cannot easily know or control the inner emotional feelings of participants. The experimental situation has its own demand characteristics. As a researcher I enter the lab with certain expectations and I might inadvertently communicate those to participants and so influence the behaviour of the participants. The participants also know it is an artificial situation and that they are being observed, so that will affect the way they behave. Critics of experimental design fear that in treating influences of the real world as variables, and in the attempt to control them, it might mean removing the very contexts that construct the behaviour we are hoping to observe.

Ethical issues are problematic in the experimental design for my study. The whole concept of experimental design reduces people to numbered subjects and the random allocation of people rather negates any value of individual differences, even though in this study, individual differences in experience and expression of emotion are of interest. In addition this study involves a potential element of deception as two groups, comprising half the participants, are misled. Within the two misled groups, one group are told the system might respond to their emotional expressions, but it won't respond, while the other group are not told that the system might act affectively, but it does act affectively. In order to control the experiment as evenly as possible in all other regards the initial information given to groups must be the same. I must consider the best and most ethical way of handling this issue (see this chapter 3.3.3 and in Chapter 4, sections 4.2.3 and 4.2.7).

3.2.4. Advantages

Advantages of the experimental method are that it is the best means by which cause and effect can be clearly established, with precise control of variables. Because conditions are controlled, the experiment can be replicated. A single experiment, particularly if numbers of participants are low, will not give definitive results so it is important that findings can be strengthened by repeating or replicating the experiment. The more often it is repeated, with the same results obtained, the more confidant we can be that the theory being tested and the results are valid. We can form a hypothesis, test it and then make another iteration with a variation of the hypothesis. We can publish details of our methods and results and subject our work to peer review. Controlling the design of the experiment means that we can randomly allocate people to groups and in this is a strength as we are less likely to have groups that differ from each other and more likely to end up with normative data. Although we are concerned with measures that are derived from qualitative methods, by using an experimental design of the experiment we add a quantitative aspect, because we can analyse the resulting data using inferential statistical tests that give quantitative results. These allow us to make assertions about the probability that the results are valid and could not have occurred through chance alone.

3.2.5. Conclusions on Experimental Approach

There would be no point in developing affective computing systems capable of recognising the emotional signals of users and responding to them in some way, unless such systems offered some sort of advantage. The benefits might be that users do better when using such systems, that they find them easier to use, that they enjoy using them more, or any or all of these. Using the experimental method to establish cause and effect, fits to my focus on the user perspective and allows exploration of these issues. The dependent variables that I seek to understand are the way users perform with affective systems, how they behave and express their emotions, how they adapt those behaviours (or don't adapt them) and how they feel about using affective system. I need to use both objective and subjective means to measure these dependent variables because there may be differences between how people actually perform and how they say they do. Because people are different and I cannot control all individual differences it is important to be aware of individual differences and measure or monitor those I cannot control and to consider any implications of these.

By controlling the majority of variables, observing, measuring or at least monitoring others – and by only varying two of them as independent variables (the information the participant has about the system and the way the system responds to them) during interaction with affective systems, I can gather data about the dependent variables (how users perform with affective systems, how they behave and express their emotions and adapt behaviours (or don't) and how they feel about using emotion in computing). This allows me to draw conclusions with far more certainty than non-experimental methods. If the independent variables are the only things that are changed then they must be responsible for any change in the dependent variables. By having sufficient numbers of participants taking part, it is possible to carry out statistical analysis of resulting data, which gives more strength to findings and more compelling arguments from results. An experimental design is the only method that allows testing of my hypotheses and can establish causation, as long as proper protocols are followed. It is important to investigate causal links given my research objectives:

1. In order to establish the extent to which people will naturally express emotions when they know they are interacting with an emotion-detecting computer, I

need to control, observe, measure or monitor other variables and vary whether people know that they are interacting with an emotion detecting computer and then to observe and measure any changes in observable behaviour to see if this variable causes changes in how participants naturally express emotions.

2. In order to establish the conditions under which the application of emotion detection can lead to improvements in subjective and/or objective measures of system usability I need to vary whether the application detects emotion and use both subjective and objective measures that inform whether changing this variable causes changes in system usability.
3. In order to establish if and to what extent participants are willing to adapt their behaviours during interaction with an apparently affective system (achieved via use of WOZ method), I need to vary whether the system is affective and observe and measure changes in behaviours to see if these are caused by that variable.

The two independent variables relating to my research questions are whether the system is said to be affective or not, (that will alter the participants' appraisal of the system); and how the system behaves, (with or without affective responses), that will be controlled via Wizard of Oz (WOZ) scenario. WOZ is explained in following section 3.3.

The dependent variables will be how people perform on the task, how their observed expression of emotion changes, and how they report on their experience of using the system.

My literature review identified a number of other variables that I will need to control, measure, or at least observe. I will control or observe other confounding variables by various means. My participants will be English-speaking students falling within a certain age group so that the variables of language and age will be eliminated. I will control for gender as far as possible by recruiting even numbers of males and females and controlling their allocation to groups by using blind randomisation to ensure equal numbers of males and females. This will eliminate selection bias; balance the groups and form a good basis for statistical tests. I will collect some demographic information about each participant and single questions will allow me to capture and observe information about cultural background and first language used etc; a Likert scale will allow me to observe the variable of self-reported experience with computers; personality type, susceptibility to affect and communication style will be observed by use of brief standardised scales; a checklist for the researcher to complete will allow me to appraise observed communication style, against the participants' self reported communication style scales (see Appendix E).

By using more than one observer and checking for inter-rater reliability I can make the observational data as rigorous as possible. The time spent prior to the experiment in administering scales also serves the purpose of allowing me to 'tune in' to individual differences of the participants, so that I am better able to observe other variables such as their physiological state, perceptual and cognitive capabilities, temporal traits, social positioning and emotional and communication skills. It is important to be aware of these dependent variables because they could confound my findings. For example for some cultures or for some individuals it is not considered polite to make an outward show of feelings and so if I am able to identify someone that falls into this category, and if they demonstrate less observable emotional behaviours during the experiment in comparison to others, I might speculate that the difference is at least modified by their individual difference, rather than by the experimental condition.

Additional questions administered following the experiment allow me to capture the participants' self reported experience of usability and emotional experience under the different conditions (See appendix E for complete set of questions and observational scales administered before and after the experiment).

Performance will be measured by simple metrics of tasks completed, and following the experiment I will use checklists and heuristics to establish participants' views on affective systems and on general usability of the experimental system.

3.3. Wizard of Oz Method

3.3.1. Method

Wizard of Oz (WOZ) is a research method in which subjects are observed while they interact with a prototype computer system in real time, that they believe to be autonomous, but which is actually being operated or partially operated by a hidden operator who processes input from the user and simulates the system output accordingly. During the evaluation, users are led to believe that they are interacting with and directly controlling the system. Users are unaware that the system was not real, at least until after the experiment. This method takes its name from Frank Baum's children's book "The Wizard of Oz" (1900) and well known as a feature film released in 1939 in which an ordinary man hides behind a curtain and pretends, through the use of smoke and mirrors, to be a powerful wizard and in this manner controls the inhabitants of the magical land of Oz and convinces his subjects that he can grant wishes.

In the development of computing, user centred, iterative design and using the WOZ method was found to be particularly useful in early prototype systems, where the researchers controlled the experimental setting in order to explore the potential or drawbacks of systems (e.g. Kelley, 1984; Gould, 1985; Green & Wei-Haas, 1985; Dahlbäck & Jönsson, 1989, 1993). In the Dahlbäck & Jönsson studies WOZ systems were configured so that participants believed the system was recognising and responding to their speech, when in reality a human intermediary was recognising their speech, and controlling the system to respond appropriately to their spoken words. The requirements of my study is similar, in that I want to study how users feel about and behave with systems that can recognise and respond to emotional expressions, but prototype systems are not yet good enough to give the evidence and answers I seek.

3.3.2. Advantages

Advantages of this method are that a working affective computing system is not required. We are not yet able to build comprehensive and capable affective systems and even to partially implement such a system would be very costly of time and effort. By using WOZ techniques I can simulate the way future systems might work and use that to gather information about the nature of interaction and to test such a system out and to identify potential benefits or issues while still in the early stages of the design process. This enables me to see what input techniques and sensing mechanisms work best and make recommendations for system design or alterations, without having to go to the expense in time and resources of actually building working systems. Because I can make it appear as if the system actually interacts affectively with the participant, the WOZ technique should give unique insights into how users behave and feel about affective systems.

3.3.3. Disadvantages

Disadvantages are that it is quite difficult to design and to maintain a WOZ scenario that is convincing, and participants may realise it is a pretence. It is important to pilot the scenario, and to ensure that conditions remain the same throughout interactions with different participants. The wizard often has to work very hard, controlling all the elements, making consistent observations of behaviours of interest and making quick but consistent responses to those behaviors as prescribed by the imagined prototype system set up, all carried out in real time. As it is a pretence this adds to the ethical dilemmas around deception of participants already identified as potentially problematic in our experimental design. Careful piloting of the experiment and design of briefings and debriefings and consent procedures are required to mitigate this (see these described in Chapter 4 and e.g.s in Appendix E).

3.3.4. Conclusion on Wizard of Oz Method

A major obstacle to my work is the lack of a realistic and reliable affective system. My review of the literature showed that while there is great expectation of future systems, in practice, current applications are very limited and do not claim to be fully functioning. One way to manage this, which particularly lends itself to experimental design in a laboratory setting, is the WOZ interactional observation method.

3.4. Observational Methods

3.4.1. Methods

As I have established that human communication of emotional information is complex and multi modal and as automatic detection of emotional expressions is in its infancy and not yet capable of this or reliable, I will have to depend on human observation of my participants. Observation must be as systematic as possible. It will involve noting and recording behaviours and interactions of interest as they occur, maintaining as far as possible a professional distance and resisting personal involvement, so that observations made are as unbiased and objective as possible.

Systematic observational methods involve choosing the situations for observation – in the case of this study - the multimodal, often non-verbal expressions connected to emotion that are used when interacting with an (apparently) affective system. After introducing the participant to the setting, observers must record behaviours of interest, trying to capture an insider view of the setting while remaining as unobtrusive as possible, by making notes and using audio-video equipment and recording observations either in real time during the interaction or as soon as possible after it.

The resulting data must be collected, interpreted and compared. Analysis of the data will involve identifying discrete behaviours or sequences of behaviours. Conversation and discourse analysis (Hutchby & Wooffitt, 1988; Ten Have, 1999; Brown & Yule, 1983) will influence my analysis of behaviours. These methods can be used to analyse naturally occurring talk and conversation and offer insights into systems of social meaning and the methods used for producing orderly social interaction. They are a useful technique for evaluating conversational interaction and can include the analysis of non verbal as well as verbal behaviour. In my study, I will extend their use, for observation and analysis of emotional expressions.

3.4.2. Disadvantages

Observational methods can be very time consuming. To make observations more rigorous and ensure they are reliable and valid requires careful preparation so that the observer is well equipped with both know how and equipment needed to record interactions. Ideally more than one researcher should check observations, although this is costly in time and effort. This is particularly so with retrospective observations that require constant running through episodes. There is a danger that repeated and detailed run-throughs may distort the observer's perception of data.

Video footage of interactions is helpful, but use of video is notoriously prone to slips and errors, so very careful controls are needed. It requires training and acquisition of skill and know-how and considerable experience to be a trained, independent, self-aware observer and to make observations efficiently. The experimental design requires the researcher to absorb and accurately record the behaviour of participants at the same time Because the observations are so embedded in a particular context, generalisations from one situation to another may be hard to draw. It is often difficult to remain objective, especially when participants do show emotional expressions and behave in emotional ways.

Ideally observation should be a non-intrusive technique, and the researcher should be both immersed in the setting they are observing and intensely aware of it and at the same time like a 'fly on the wall' so that they do not intrude into the setting or create an impact on it and the participants should not be aware of observations. This is a fine line to tread and in experimental set ups the researcher usually has to introduce the experiment to participants and negotiate their participation. It may take time for the researcher to 'melt into the background' and it may be difficult to ensure that participants behave in a normal way, so there is a danger of a Hawthorne effect (Gillespie, 1991) where participants change the way they behave, simply because they are motivated by the interest shown in them from being observed. A Pygmalion (or Rosenthal) effect might occur, where the researcher communicates their expectations to the participant or a placebo effect due to the participants' perceptions that they are receiving an intervention, even if they are not (Rosenthal & Jacobson, 1992). The observational method generates a lot of data that requires processing and analysis. And so methods are needed to manage this issue.

3.4.3. Advantages

By using observational methods it is possible to record spontaneous behaviour of participants. Ethnographic approaches can be incorporated into observational methods, to include wider understandings of behaviours and how they are influenced by understandings, beliefs and attitudes. A major strength of themed analysis or Grounded Theory is that they can capture unexpected data that might otherwise be missed (Glaser, 1998). Categories of data can be defined beforehand, but can also or instead be identified from what is there – so that the theory emerges from the data on the ground rather than from pre-defined theories that might narrow the findings.

Observation of behaviours can include interactions with features and artefacts and can include noticing norms, values, procedures and rituals and their physical, social and cultural settings. Observational techniques can result in in-depth information with rich insights into participants; their understanding of situations; their resulting behaviours (including their non verbal behaviours) and their whole world of meaning. Valuable evidence about the outcomes of the intervention can be noted along with effects of the intervention and the social context or setting of the intervention. Any real time observations can be validated by video recordings that allow retrospective observations and analysis to be made by the original observer and validation by other researchers. There are tried and tested methods such as structured observational frameworks, sampling of time sequences, and ways to develop coding systems (Bakeman & Gottman, 1997; Bakeman & Quera, 1995) that can help with observational methods so that what is essentially qualitative data can be quantified, aggregated and statistically and sequentially analysed for generalisations to be made.

3.4.4. Conclusions on Observational Methods

Observational methods are particularly well suited for this study, as I am planning an intervention that is innovative, with a number of confounding factors that might influence success or failure. Observational methods will enable user focus and allow me to collect rich data and gain insights. By starting without too many preconceptions I am open to whatever range of behaviours occurs and by using a themed analysis as part of the method, I can build on theory as I progress.

3.5. Subjective Measurement Methods

3.5.1. Methods

There are two main methods for measuring emotional feelings, behaviours and expressions that are particularly used in studies of affective computing: categorised as quantitative or qualitative. Researchers are generally divided into those measuring quantitative physiological signals such as muscle activation, galvanic skin response, heart rate etc and those measuring qualitative observed expressions or self reported feelings.

3.5.2. Quantitative Methods

We are able to detect and measure more and more signals with reliability, and more and more can be automatically detected using machine recognition, well summarised by the relevant Humaine deliverable, (Gelin, 2005). Even when reliably detected, *interpretation* of these signals and the context in which they are generated is more problematic.

It is debatable whether it is possible to use quantitative methods to assess inner feelings. There is some evidence that certain physiological measures are related to emotion. For example Hazlett used facial electromyography (EMG) to assess emotional states (Hazlett, 2003; Benedek & Hazlett, 2005) and it is generally accepted that galvanic skin response is increased with high arousal, although it is not possible to specify the nature or valence of the emotion that is associated with it. Galvanic skin response can increase for other reasons, such as a rise in body temperature, so other causes of high readings include sweating due to illness (e.g. due to flu), or due to getting hot during physical exertion (e.g. when running). Comparisons between individuals is difficult because of underlying differences between individuals, so the baseline of what is normal for one person in one situation might be higher or lower than for another individual.

There is growing evidence that some physiological changes can be reliably correlated with certain emotional states in certain circumstances. For example Levenson, Ekman and Friesen use a directed facial action task where subjects re-lived or imagined an emotion task (1990) and they found that heart rate and finger temperature changed to reliably show fear, anger, disgust and happiness.

Gaming is an area where there has been a lot of interest in physiological measures. Possibly as gamers are usually static and 'wired' by a controller in any case, the use of wires and sensors is more acceptable or plausible. Arousal (which can be relatively easily measured) is of particular interest in gaming. Mandryk, Inkpen and Calvert (2006) suggest that physiological measures follow trends of reported subjective experiences in a game environment and also differentiated between different play conditions. Later development of this work suggests modification of play dependent on physiological state (Mandryk & Atkins, 2007).

Undoubtedly physiological measures are of interest, and likely to play an important role in future systems, but they do have severe limitations at present. Firstly, we simply do not understand enough about how physiological signals are linked to emotion, particularly at higher levels of function (e.g. cognitive, reflective) or in varying interactional or social contexts. One of the areas where there is potential for affective systems is with groups that have neurological or physiological differences (e.g. autism) and they might not present with typical activity patterns (Müller, Kleinhans, Kemmotsu, Pierce & Courchesne**,** 2003; Silani, Bird, Brindley, Singer, Frith & Frith, 2008). Secondly, current systems are rather intrusive in nature, as subjects are often required to wear sensors and are attached to measurement devices with wires during data collection, which restrict movement and arguably might themselves change the nature of affective experience e.g. depression monitoring by a head worn device (Hu, Moore, & Wan, 2008).

As studies of the brain and associated emotional states and physiology lead to better understanding there may be greater benefits from using physiological measures. Less intrusive systems are under development and remote or wearable wireless devices to collect physiological information are becoming more feasible and reliable, such as remote sensors to collect heart readings (Harland, Clark & Prance, 2002), Peter's affective glove (Peter, Ebert & Beikirch, 2005); wristbands that pick up galvanic skin measures (Poh, Swenson & Picard, 2009); and cameras or wearable devices to detect and analyse movement patterns, e.g. el Kaliouby and Robinson's (2004) system for 'face-reading' based on facial movements and Baron-Cohen's Mind Reading taxonomy (Baron-Cohen, 2003) and wearable Lillypad technologies (e.g. Fajardo & Moere, 2008). In addition physiological detection of features is now approaching real time delivery of analysis of certain features (e.g. el Kaliouby & Robinson, 2004).

3.5.3. Qualitative Methods

Undoubtedly human appraisal of emotional states using observational or self report measures is unrivalled by machine detection at present, in that it is rich and complex and situated in real contexts. But we know that humans vary in their abilities to perceive and interpret emotional states or specific attributes of them. Critics suggest relying on such methods is subjective and liable to bias.

Feeling is by its nature subjective and it could be argued that in some ways it can be best assessed through asking participants to explain how they feel. However there are many complications due to different individuals having different inner experiences; having different meanings associated with emotion words; or different understandings of behaviours or language. Just thinking about what you are feeling may alter that feeling. It is also difficult for people to make comparisons with feelings or perceptions of feelings of others.

Expressive behaviours should lend themselves to identification and measurement by qualitative methods. In practice this is very complex, as humans naturally take in a whole range of simultaneous multimodal messages. Even from a quick glance or listening to a single word a person will assess another and start to make inferences from their appearance, where they are situated, the timing of their stance, their accent or dialect, pitch, tone of voice and so on. They will immediately infer factual information such as gender, age, educational status and emotional status as well as inferences about more abstract factors such as mood, attitude and stance. In contrast to naturalistic attention to multiple modes, coding of expressive behaviours often requires attention to a single mode, for example, attention to facial expression in order to code all instances of smiling. It is somewhat unnatural to attend only to one mode and doing so may completely alter awareness of emotional expression. If only one mode, or a restricted number of modes are coded, then other modes that contribute information about emotional expressions may be missed. Sometimes emotional information is carried in the timing of modes in conjunction with other modes, for example saying something in a sad tone, but accompanying it with a smile or following it with a wink of the eye, might indicate that it was a 'mock' sadness. If only the smile is coded and not spoken words, intonation or other behaviours, occurring synchronously or asynchronously then the true expression would be missed.

This can be partly addressed by applying systematic qualitative methods of evaluating behaviours. There are a number of questionnaires used as measures of certain dimensions of affect that have been rigorously developed and validated e.g there are hundreds of standardised scales for measuring anxiety and depression (e.g. Hedlund & Viewig, 1979; Beck, 2006; Taylor, 1953). However these are not really suited to interactive contexts as they tend to be lengthy and can only be used prospectively or retrospectively to a task. They may be unnecessarily long, as in the case of depression scales, where a single carefully worded question such as "Do you frequently feel sad or depressed?" has been found just as valid as longer questionnaires (Watkins, Daniels, Jack, Dickinson, & van den Broek, 2001). Some studies have used 'probes' where users are asked to state their felt emotions at certain points in time during interaction with an application. However this breaks the flow of engagement and possibly a sudden need for reflection might alter the experience of emotion. Russell's Circumplex is frequently used as a basis for measurement, for example as the basis for Bradley and Lang's (1994) Self Assessment Manikin, that allows users to rate their emotional intensity and valence on a Likert scale by indicating which small cartoon manikin best relates to their felt state.

Self-report measures often rely on users indicating words they feel are relevant to their emotional states from lists that are constructed from various taxonomies. These can yield some interesting data but we have no way of knowing their exact interpretation of words used. For example 'happy' might mean different things to different people. Silzer describes the difficulties in translating emotion words such as 'Miffed, Upset, Angry, or Furious' (Brown & Silzer, 2001).

Questionnaires and scales often relate to particular states such as depression, motivation, anxiety or susceptibility to affect scales (e.g. see Riek, Afzal, & Robinson, 2008). These are often intended for specific uses e.g. in health care or psychology settings for management of clinical depression. These give detailed breakdowns of factors relevant to each scale but tend to be quite a burden for participants to complete and time consuming to administer and score.

Methods from other more established domains offer ideas for systematic and less subjective methods for evaluation of affect in affective computing systems. Observational data can be used to establish relevant behaviours that can then be measured and analysed. Bakeman and Gottman (1997) and Bakeman and Quera (1995) have developed interesting and successful strategies for sequential analysis of interaction. Rather than observing models and fitting observation schemes to those models, their work involves observing interactions, defining mutually exhaustive and exclusive factors, and devising coding protocols to capture those factors and then building models to fit what is observed. For example, the Specific Affect Coding System (SPAFF) was developed to assess marital interaction, and uses five positive codes such as interest or humour and ten negative codes, such as disgust or anger. The foundation of this method is the development of a catalogue of behaviour codes relevant to the particular project. Bakeman and Gottman describe these as 'the measuring instruments of observational research; they specify which behaviour is to be selected from the passing stream and recorded for subsequent study' (1997, p. 4). They found that having a coding strategy using levels of coding gave more 'hooks for observers to grab', leading to collection of more reliable data. Ethnographic studies and linguistic anthropology use discourse analysis methods with careful coding of behaviours and themed analysis or Grounded Theory to analyse behaviours, in context, including those related to affect. For example Charles Goodwin's (2004) analysis of non verbal interaction with an individual who has had a stroke or Elinor Ochs' studies of middle class families in California (Ochs, Graesch, Mittman, Bradbury & Repetti, 2006).

Systemic functional linguistics is a fast growing field that seeks methods to understand the messages we send in both what we say and how we say it. The choice of words and the attitudinal stance that they convey can be analysed in a systematic and detailed manner, for example using the appraisal framework (Martin & White, 2005):

> ".....an approach to exploring, describing and explaining the way language is used to evaluate, to adopt stances, to construct textual personas and to manage interpersonal positioning and relationships."

To date this approach has mostly been used with written language, for example see text analysis work on Appraisal Framework website (Martin & White, 2005), which is gradually being extended to include visual representation in images. It has potential to extend to non-verbal interaction analysis.

Researchers often use video coding of behaviours, performed manually, which is a very time intensive method (sometimes many hours of coding for a video segment that lasts only seconds). This involves making videos of people during interactions and then going through these and marking and tagging episodes as relating to particular emotion related behaviours of interest. For example using videos to identify when people nod their heads and what this seems to mean (Lee, Lesh, Sidner, Morency, Kapoor, & Darrell, 2004). With improvements in automatic recognition systems it is sometimes possible to reliably and automatically identify certain behaviours, for example Facial Expression Analysis Tool (FEAT) (Kaiser & Wherle, 1992) is a system to automatically detect movements of certain groups of facial muscles, based on Ekman's Facial Coding System (Ekman & Friesen, 1978). Automatic detection of behaviours can save a great deal of time in coding, but there are drawbacks. Firstly there may be associated behaviours of importance that you are not aware of, and secondly implication of a one to one correlation between actions and meanings. For example a person might be telling a joke requiring a mock expression of worry. The system might automatically detect a group of facial actions (e.g. the frowning) and code it as 'worry', but unless it also accesses the context of the spoken word it will be a misinterpretation. Or the person might be using gaze in conjunction with the frown to indicate another person (as the subject or cause of his worry) and that might be missed if context is not considered. Context recognition is in its infancy but is beginning to be addressed, for example the probability of states is predicted (e.g. Morency, deKok, & Gratch, 2008).

Another use of human coding is to check or act as data to instruct automated systems. For example Gunes and Piccardi (2006) used an annotated corpus of data, coded by human coders for face and body movements, to train an automated system.

3.5.4. Combined Methods

There is an interesting move to incorporate objective and subjective measures. Leahu, Schwenk & Sengers (2008) suggest the need to balance subjective and objective measures and they allow subjective interpretation of objective physiological signals using bio-mapping and other techniques. Nold (2009) uses bio-mapping by combining galvanic skin reading with geographic position (established via GPS positioning technologies), so that an 'emotion map' of a journey can be drawn. Boehner, Sengers, & Warner (2008) suggest that assessing only well defined and testable precise representations may mean missing the fullness of the experienced phenomena and they developed prototype systems that encourage users to view representations of physiological states and make their own interpretations.

Natural Language Processing (NLP) involves computer-based analysis and generation of language by applying various algorithms and specialist methods. This is proving useful in analysis of how people use emotion related words and there are now attempts to combine some paralinguistic features, for example taking physiological information about tone of voice and synthesising that with the meaning of the spoken word (Schuller, Muller, Lang & Rigoll, 2005).

3.5.5. Conclusions on Subjective Measurement Methods

There are advantages and disadvantages to consider regarding methods for assessing emotions. A wider inter-disciplinary perspective will be useful here. There are a number of ways of measuring both inner feelings and outward expressions of emotion.

Quantitative measures are very much a ‘work in progress’ and are at present rather intrusive, although particularly useful in certain contexts. They are not yet reliable or comprehensive, especially in settings where context is of interest. Inner feelings can sometimes be gauged by physiological signals, although the measurement devices tend to be intrusive and exact correlations between measures and experienced feelings is not exact. They do not generally take context into account, at all.

Any use of qualitative measures will require a systematic approach to avoid subjectivity and bias as much as possible. Some methods from other domains such as social science or linguistics may be useful. Some validated scales are useful. Russell’s circumplex model of affect has proved of great practical usefulness as a basis for measurement of affect in HCI.

There are compelling arguments for both qualitative and quantitative approaches and some promising results from combinations of measures. The use of combined methods may be particularly useful in a setting where the user’s own interpretation can be brought into the system design.

In this study I will seek the best of both worlds, by following the experimental method and collecting data that can be both quantitatively and qualitatively analysed in order to give both statistically significant results and also deeper insights into user needs and design inspirations for future systems.

3.6. Tools

3.6.1. Affective Application

On a practical level, I need to design and develop an application that will act as either an affective or non affective environment and be a vehicle to test out my hypotheses. From my review of the literature, both gaming and learning environments are areas that hold promise for affective systems. An interactive word game or puzzle would provide a suitable environment and provide opportunities for both textual and visual representations that might elicit some emotional behaviours and responses.

Methods to elicit mood or affect have been used by a number of researchers ever since the seminal work by Schachter & Singer (1962) (and see reviews of mood induction procedures by Martin, 1990; Gerrards-Hesse, Spies & Hesse, 1994); Westermann, Spies, Stahl & Hesse, 1996). Velten (1968) devised validated mood statements that are associated with positive and negative affect and Lang, Öhman & Vaitl (1988) devised sets of stimuli including sounds and pictures (e.g. the International Affective Digital Sounds and International Affective Picture System) to elicit reliable affective responses. I can learn from these sets of pictures and statements, in order to inspire design of a game that will elicit emotional responses. Other features related to poor design (Ward, Marsden, Cahill and Johnson, 2002) such as time pressure or system interrupts or delayed system responses (Riseberg, Klein, Fernandez & Picard, 1997; Scheirer, Fernandez, Picard & Klein, 2002) can be added to offer an element of frustration for the user. ‘Word ladders’, reputedly invented by Lewis Caroll, (Gardner, 1996) are a good example of a simple word game that can use both visual and verbal clues.

3.6.2. Range of Coding and Analysis Tools

Coding schemes There are a number of coding schemes and tools to help organise observations of behaviour. These have been developed over the years for different reasons, and fit particularly well to certain fields. These have been comprehensively surveyed (e.g. Humaine toolbox, 2009; Knudsen, Martin, Dybkjær, Berman, Bernsen, Choukri, Heid, Kita, Mapelli, Pelachaud, Poggi, van Elswijk & Wittenburg, 2002; Bird & Harrington, 2009). At the time of this experimental work the use of Facial Action Coding scheme (FACS) was much to the fore. Based on the work of Ekman (Ekman & Friesen, 1978, Ekman, 1982) this facilitates manual coding of video on a frame by frame basis. Trained observers look for 44 ‘action units’ that are movements of particular groups of facial muscles. In other domains different labelling schemes have been developed for particular types of movement or movements of particular parts of the body, for example Labanotation (see Hutchinson-Guest, 1989) is used in choreography to plot dance movements and has been applied to body movements in other domains. Some automatic coding systems are being developed, for example based on the Facial Action Coding System (FACS), but again these are specific to particular aspects of emotion related behaviours. Using coding tools and a number of coding passes through the data facilitates synthesis of a number of different coding structures.

Coding tools range from use of tactile objects, such as the sensory evaluation instrument, (Isbister, Höök, Sharp & Laaksolahti, 2007), intended for naive users to show their emotional states, to complex computer applications intended for use in research analysis. Typically these applications are qualitative analysis packages that have a video window and allow portions of the video to have coding decisions added to them as metadata which is represented onscreen in some way. By reviewing the video a number of times and adding metadata for different aspects of behaviour to the same portion of video it is possible to build a comprehensive record of the behaviour observed.

During the coding exercise, it is important to control for any possible individual bias of any observer by using multiple observers to code the same data and establishing a satisfactory degree of inter-rater reliability (Lombard, Snyder-Duch, & Bracken, 2002), (see later discussion 4.2.9; 4.2.10; 7.7).

Features of various tools (e.g. reviewed by Zeng, Pantic, Roisman, & Huang, 2007) include:

- timeline bar displays as used in Transana, Anvil and Observer (Woods & Fassnacht, 2007; Kipp, 2001; Noldus Technology, 2001). Mouse or keyboard entries add time stamps to transcripts or timelines displayed with video. Multiple behaviours can be defined, separately coded and displayed as differently coloured bars;
- likert scales such as the Self Assessment Manikin (SAM) (Bradley & Lang, 1994) allow coding of the scale of a behavior (e.g positive to negative, valence or arousal). Scales are completed on paper or screen retrospectively. Resulting codes are not linked to video or timeline. This tool has been used to explore

responses to Velten (1968) mood statements (Jennings, McGinnis, Lovejoy, & Stirling, 2000);

- two dimensions, colour and real-time representation are used in Feeltrace (Cowie, Douglas-Cowie, Savvidou, McMahon, Sawey, & Schrööder, 2000). Input is via a mouse. Time line display shows activation on the y axis with colour shading under the curve to show emotion type;
- six simultaneous curves in real time show probability of different complex states in Facereader (Noldus, 2009; el Kaliouby & Robinson, 2004) using machine recognition to infer mental states and display them as colour coded probabilistic curves on a timeline;
- texture and colour conceptualise emotional states in the Sensory Evaluation Instrument and eMoto (Isbister, Höök, Sharp & Laaksolahti, 2007; Fagerberg, Ståhl & Höök, 2004);
- the Continual Measurement System (CMS) (Messinger, 2009) is designed to facilitate continual measurement of affect by expert or non-expert coders using a joy stick or a mouse.

Qualitative analysis packages (as reviewed by Lewins & Silver, 2007) can help with management of transcripts and include e.g. NVivo or AtlasTi (Nvivo, 2009; Atlas Ti, 2009) or for applications that can import video data such as Transana, Anvil or Observer (Woods & Fassnacht, 2007; Kipp, 2001; Noldus Technology, 2009). Latest generation computer assisted qualitative data analysis (CAQDAS) tools such as MiMeg (Tutt, Shaukat, Fraser, Hindmarsh & McCarthy, 2007) and the Digital Replay System (Brundell, Tennent, Greenhalgh, Knight, Crabtree, O'Malley, Ainsworth, Clarke, Carter & Adolphs, 2008) allow sharing of data and integration of signals from continuous monitoring devices, so that for the first time we can visualise both qualitative coded labels and quantitative physiological or other continuous data streams such as global positioning system (GPS) data at the same time. These allow instances of behaviours to be coded, labelled or tagged and assist a grounded theory approach or themed analysis, so that 'trees' or 'families' of behaviours can be built up and arranged as a series of video clips. This allows review and reflection of groups of clips to check that they are correctly labelled, to go over them in more detail or to count or otherwise analyse them. Coding decisions can be checked and validated with other researchers or with participants to make sure that interpretations of behaviours have been correctly made or to see comments and reactions.

Outputs from these can be used in statistical packages such as SPSS or GSEQ (Bakeman & Quera, 1995) and some applications include or link to sophisticated sequential analysis and data mining tools, but are expensive e.g. Observer and Theme applications (Noldus) (Noldus Technology, 2009).

3.6.3. Conclusions on Tools

An interactive word game with textual and visual cues would provide a suitable environment for my experiment. Word ladders are a very well established word game that could be easily developed into an interactive game with both visual and verbal clues for words, which along with some added features to invoke frustration such as added time pressure, system interrupts and delays, would afford opportunities to elicit both positive and negative states from participants.

There are a number of coding methods and tools available and more emerging. Coding schemes tend to be restricted to particular types of movement, so if using these, repeated measures with a range of schemes must be used to capture all aspects of emotional behaviour. Each tool to assist with coding practice has some useful features, although none have all the features in the one tool. There are a number of qualitative analysis packages available to aid in the complexities of collecting observational data, organising, measuring and analysing it, and a range of statistical packages to help with analysis.

3.7. Summary

Observational methods are particularly well suited for my study and will allow me to focus on and gain insights into the user perspective. Observational methods allow me to remain open to any unexpected ways that emotions are outwardly expressed by both verbal and non-verbal means and to consider the context in which they take

place. I will use a systematic approach to observations to avoid subjectivity and bias as much as possible.
There are a number of ways of measuring both inner feelings and outward expressions of emotion and I will use subjective qualitative coding measures, because they are less intrusive and because quantitative physiological measures are not yet reliable and have only limited known correlations to emotional states.

By developing an interactive game, designed to elicit some emotional behaviours, I can provide an affective or non affective context for my experiment. Coding and analysis tools will support my collection, organisation, measurement and analysis of observational data and enable me to ensure reliability of my methods, for example by ensuring that inter-rater agreement is satisfactory.

Incorporating other methods will support observational methods, including:

- methods from psychology such as validated scales and user questionnaires;
- methods borrowed from psychology and established in HCI, such as categorising emotions using the circumplex model and the self assessment manikin;
- methods from social science and other domains such as time sampling and behaviour coding;
- methods from linguistics such as discourse analysis.

My study will use an experimental design with independent, dependent and controlled and observed variables, in order to control the variables and create a replicable study that provides empirical evidence, which is needed in this domain. Ethical issues need and will receive careful management.

In my study, in order to minimise the burden on participants and reduce administration time, while still including some well validated and reliable measures I will use a limited number and short versions of scales to observe some variables such as susceptibility to affect, communication style and personality type and to collect some data such as individual attitude to affective computing and self reporting of inner feelings about emotion.

Sets of questions and scales included in this study are:

- Self reported demographic information e.g. data on age, culture. languages spoken, family socio-economic status, experience of and attitude to using computers;
- Communication style questionnaire (Stephen & Harrison, 1986) and communication style observation schedule to establish if habitual communication style is within norms, and if individual has a tendency to aggressive, passive or assertive style;
- Self Assessment Manikin (SAM) (Bradley & Lang, 1994) for self report of emotional valence and intensity;
- Positive and Negative Affect Scale (PANAS) for self report of affective state, (Watson, Clark & Tellegen, 1988);
- Usability heuristics - set of 10 questions (based on Nielsen, 1994, 2009) for self report of usability of different conditions (and also included in order to give credence to the briefing given to participants that leads them to believe the focus is on usability of the system, rather than their emotional state);
- Affect Intensity Scale (short form) (Geuens & de Pelsmacker, 2002) for check on individual susceptibility to positive and negative affect;
- Big 5 personality test (Goldberg, 1992) to assess personality style;
- Set of questions relating to experience of the participant, and their views of whether they showed emotions, their attitude to the concept of affective machines, etc.

By using the Wizard of Oz method and experimental design, along with various tools I can carry out interactional observations and compare groups of participants acting under different conditions in a laboratory setting despite lacking of a truly affective system, but this will need careful design and management. Chapter Four goes on to show in detail how I have used these methods in my experimental study.

4. Affective Experiment Design

4.1. Introduction

This chapter describes the experimental design used for this study and illustrates the complex considerations required when working with emotion in computing. It details the work involved in designing and developing the overall experiment including:

- development of experimental hypotheses (4.2.1);
- the 2x2 factor experimental design (4.2.2);
- the dependent variables (4.2.3);
- the recruitment of participants and design of briefings of them, to address ethical concerns (4.2.4);
- the design and development of a game to provoke some emotional responses and behaviours from the participants (4.2.5);
- the equipment required (4.2.6);
- the setting of the experiment and development of a convincing WOZ scenario and protocols for procedure (4.2.7);
- the development of the plan for general analysis (4.2.8);
- the development of a coding procedure (4.2.9);
- the work on inter-coder reliability (4.2.10).

This study design illustrates the very wide considerations that have to be taken into account in this field to successfully simulate an affective system and to provoke and understand the very rich emotional interactional behaviours of participants.

My review of the literature in Chapter Two established that user-centred studies are needed to inform development of affective systems. In particular I need to establish if and how people show emotions during interaction with affective systems. I need to establish if people gain in terms of performance or satisfaction when using affective systems. Once a method to establish these basics is formulated, then different types of adaptive systems and adaptive user behaviours can begin to be explored. I found that the current state of development of systems that use emotions are very limited, tend to be technologically driven, use restricted definitions of emotion and focus on only some aspects of emotion, with various levels of system response and with little exploration (so far) of ethical issues arising from use of such technologies. This is partly because despite technological advances, systems do not yet achieve the multimodal capacities of people.

Following my critique of methods in Chapter Three, I decided this could be overcome by simulating capabilities of future systems by using a Wizard of Oz method and by using experimental design to control an affective system, so that it appears to respond emotionally in a 'normal' human way to participants.

This chapter describes my experimental study design in detail, showing how by employing a two by two factor experimental design I am able to control and compare four groups of participants. In the experiment described here, the independent variables are the information given to the participants and the way the system appears to act. This means that different groups of participants operate under different conditions, because they have different contexts and are given different information:

- firstly context varies as to whether individuals are interacting with an affective system or not (one that does or does not appear to respond to their affective communication);
- secondly information varies as to whether individuals are told the system is affective or not (so that they do or do not expect the system to react to their emotional behaviours).

This chapter goes on to explain how this study uses observational methods and subjective measurements to record and analyse any variance in the resulting behaviour of users and changes in their use of emotional expressions. It describes in detail participant recruitment, the development of the game designed for use in the WOZ scenario, the equipment to capture data and strategies used to make the overall scenario convincing. Finally, the rationale for the coding procedures and methods and the reliability checks used as part of the analysis are explained.

4.2. Study Design

This study uses experimental design with a multi-method approach involving methods for the social sciences (particularly in my evaluation of behaviours) in order to address the many challenges of working in the affective computing arena identified in my exploration of literature and methods relating to affective computing (in Chapters Two and Three). In the literature searches and pilot work I identified a number of potential confounding variables to control (if reasonable), or if not, to observe and consider in the results, as explained in the methods chapter. For example I will eliminate age differences by limiting participants to a certain age range and make gender a constant by ensuring an even distribution of genders between groups.

4.2.1. Hypotheses

The objectives of the research identified from my review of the literature in Chapter Two are:

1) Relating to observable behaviour under different knowledge conditions
- to establish the extent to which people will naturally express emotions when they know they are interacting with an emotion-detecting computer;

2) Relating to task performance and user satisfaction under different conditions
- to identify the conditions under which the application of emotion detection can lead to improvements in subjective and/or objective measures of system usability;

3) Relating to observable behaviour under different conditions
- to establish if and to what extent participants are willing to adapt their behaviours during interaction with an apparently affective system.

The main experiment in this project, as described in this chapter, is designed to assess if there are any improvements in performance or in system usability. It explores the extent to which participants display emotion at the interface, either:

1.under normal conditions;

2.when a system appears to respond to their emotional displays;

or

3.when they believe the system can recognise and respond to their expressed emotions.

Arising from the general objectives are specific research questions:

1. What is the effect of interacting with a system that acts affectively on expressions of emotion?
2. What is the effect of interacting with a system that acts affectively on adaptation of user behaviours?
3. What is the effect of a system that acts affectively on task performance of users?
4. What is the effect of a system that acts affectively on reported user satisfaction?

If Reeves and Nass (1996) 'Media Equation' is to hold, and the participants treat the computer system like a real person, then it is to be expected that participants' performance, satisfaction and emotion related behaviours will depend on their experience of the actual behaviour of the system. If the system acts like a real person, by responding sympathetically to emotional displays, then people will be likely to perform better, like the system more and respond in kind by increasing their emotional displays. Therefore the hypotheses developed from the research questions are:

H1. If the system acts affectively, there will be improved performance by the participant.

H2. If the system acts affectively there will be improved satisfaction on the part of the participant.

H3. If the system acts affectively, then participants' expressions of emotion will increase in valence.

H4. If the system acts affectively, then participants' expressions of emotion will increase in intensity.

H5. If the system acts affectively, then participants' emotion related behaviours will change.

In a number of pilot observation exercises, I explored ways to alter the appraisal of participants so that they would think the system acted affectively, and so that I could see if their subsequent behaviour and use of emotional expressions changed. I suspected that if they were given different information about the system and if I could convince them that the system could detect their emotional expressions, (even if in fact it did not act that way), then they would have different expectations, different experiences, behave differently and achieve different outcomes from it.

4.2.2. Factorial Design - independent variables

In order to satisfy the experimental requirements, I needed to inform half of the participants that they were interacting with a system that could respond to their emotional expressions (the experimental group) and half need not be so informed (the control group) (ethical issues this raises are discussed in 4.2.7). Secondly I decided to vary the behaviour of the system, so that in the case of half the participants the system responded to their affective state (the experimental group) (achieved via a Wizard of Oz manipulation) and with the other half it did not (controls). This will mean I have four groups to compare.

By comparing results between the control group (who give a baseline) and the experimental groups, I can measure the effects of the intervention. The measures I adopt for independent variables are:

1. ***Acted affective*** (with two levels; 'standard' vs. 'system appears affective'). This refers to whether the system appeared to act affectively. In the 'standard' condition clues and messages in the game appear only in response to the user clicking the 'help' button. In the 'system appears affective' condition, if the participant is observed via the one way mirror to use emotional expressions, for example looking puzzled or irritated, the game can be controlled to offer help so it appears to respond to their emotional expressions via the 'wizard'. The same 'help' features ae available to all groups, whether via the help button or via the wizard.

2. ***Told affective*** (with two levels; 'expect standard system' vs. 'expect affective system'). This refers to whether the participant expects the system to act affectively. In the 'expect standard system' condition participants are told they are testing the usability of a word game. In the 'expect affective system' condition they are told that they are testing the usability of a word game on a prototype system, which might recognise or respond to some of their emotions. There are therefore four experimental conditions in total, representing the factorial combination of the two factors.

	Application appeared to react to emotional expressions	Standard application
Participants told application was affective	Group 1	Group 2
Participants not told application was affective	Group 3	Group 4

Table 4. 2x2 factorial design

4.2.3. Dependent variables

The dependent variables I want to measure are task performance and satisfaction of users. Performance can be measured by how well participants perform when playing the game and this can be easily established by counting the number of completed 'rungs' on the ladder, for each participant. Satisfaction can be measured in a number of ways. The most important measure is how participants themselves feel about the interaction, and if they report themselves as feeling happier from it. Subjective satisfaction can be measured from the participant's own ratings of emotional valence (from happy to sad), given on the Self Assessment Manikin and /or from results from completion of the Positive and Negative Affect Scale and from qualitative data elicited by a short questionnaire. An objective view of satisfaction can be assessed by the researcher observing the interaction and analysing the participants' behaviour to see if they appear happier during interactions with the affective version of the system, compared to the standard system. (See Appendix E for pre and post experimental questions).

Other variables identified as relevant in my literature searches are eliminated, controlled or observed during this experimental research. Gender, age and language are controlled by my inclusion and exclusion criteria and the process of recruitment and allocation to groups (described in following section 4.2.4). Other individual differences that might influence the participants' use of multimodal expressions are observed, including demographic data such as, languages spoken, culture, computer experience; personal information relating to emotional communication such as mood on the day, susceptibility to affect, communication style, and personality. Data from activity logs, showing physical behaviour during game play, such as mouse movements and keystrokes made, will be collected and observed. Participants will be asked their views on usability of the system, for added data and in order to maintain the WOZ scenario which relies on them believing they are testing the usability of a system. (See Appendix E for pre and post experimental questions).

4.2.4. Participants

The design was for 60 participants with 15 per condition. They were recruited through advertisements placed around the university and by word of mouth, (see Appendix E for example of advert used). Criteria for exclusion were that they must not belong to the computing department, in case they would be likely to have heard about the study or recognise my methods and be biased by that knowledge. My inclusion criteria were that they should fall within the 18-40 age group, have normal ranges of hearing and vision, be able to give informed consent and have a good command of English (required for the word puzzle).

In practice, I had to recruit three additional participants, due to having to discard data later on, so sixty three participants took part in the study. Data from 2 subjects was discarded due to data collection failures and one due to atypical behaviour and language patterns. It proved harder to recruit females than males, resulting in 42 male and 21 female participants. These were allocated to each experimental condition. Data from the remaining 60 participants (40 male and 20 female), yielded 15 from each experimental condition, for use in analysis. All were currently living in the UK and most were graduate and undergraduate students. Participants were paid to cover expenses relating to their participation. The majority of participants were aged between 18 and 25 with age range 18-37 (mean age 21.37). I used a random allocation method (numbered squares in bags) to place 15 participants in each of the four groups with an equal distribution of males and females in each group.

4.2.5. The Game

I wanted to select a task that might lend itself to a real world affective application. Both gaming and educational settings (as discussed in sections 2.5.2 and 3.6.1) are domains where affective computing might bring enhancements. The literature search showed evidence that emotions enhance the experience of playing games, aid problem solving and improve satisfaction with decision making, so a problem solving game is a suitable application to test if affective interventions offer any benefits. The experimental task decided on for this study was a relatively simple on-screen word puzzle, called a ‘word ladder’ originally devised by Lewis Carroll (Gardner, 1996) to teach vocabulary and spelling skills. The word ladders chosen are presented on-screen and each word-ladder consists of simple four-letter words, (as shown in Figures 10, 11 and 12), involving problem solving, but consisting of relatively simple words, as in this case, the task is not intended as a test of literacy.

By choosing a problem solving game that involves pictures and words, I open the opportunity to select words and pictures likely to invoke some emotional responses. I was constrained by the words contained in each ladder, so considered many dozens of possible ladders before selecting six where it seemed possible to illustrate the words with some positive and negative statements and pictures. Many words lent themselves more to positive connotations than negative, so some alternative ways to elicit negative feelings such as frustration were considered. By adding complexity and barriers and rewards to the game I can further add to the potential for an emotional experience that involves both positive and negative aspects. I introduced a system interrupt that occurred once during the game with a pop up window that had to be closed, to delay and frustrate the user. I introduced congratulatory messages at the end of each game to reward the user. Each game was played ‘against the clock’ with a timer continually showing the amount of time remaining. In the event, due to system load, the mouse was also ‘sticky’ and I allowed this feature to remain as an added source of frustration.

A set of word ladders were trialled on a pilot group of seven people before being refined to three word ladders, each consisting of seven target words with one picture and two textual clues per word.

In the word ladder game, individuals attempt to transform the initial word into the target word by changing one letter on each of a number of given lines, or ‘rungs’, in response to clues (demonstrator in figure 11). Each word can only differ in spelling from the previous word by one letter. Participants are given the initial word for the first rung of the ladder and a target word for the last rung of the ladder and offered a series of picture and text clues to help complete the rungs in-between.

Clues are designed to invoke some emotional reactions. I explored Velten (1968) mood statements and pictures used in the International Affective Picture System (Lang, Öhman & Vaitl, 1988) for inspiration for clues that would be likely to elicit some affective responses. The resulting set of pictures and clues includes a range of images and references, including everything from cute babies to comic scenarios and funereal scenes and so they reflected both positive and negative concepts. Figure12. gives some screenshots from the application (full list of clues in Appendix E).

HEAD
BEER

Figure 10. Sample structure of word-ladder game as presented for solution

Given stucture	Given clues	Solutions
Given starter word = HEAD		HEAD
Space for solution	To listen	HEAR
Space for solution	Animal that growls	BEAR
Given end word = BEER		BEER

Figure 11. Sample structure of word-ladder game with clues and solutions

Figure 12. Screen shots of Word-ladder application

Each target word has a picture clue, a quite difficult textual clue and an easy text clue. In the standard system condition, in response to hitting the ‘help’ button, the picture clue is the default first clue to appear and the more difficult text clue is presented next. The easiest text clue is the last to appear. In the affective condition the same sequence is followed, and the participants can still use the help button as and when they wish, but the clues are also supplied in response to the participants' emotional expressions if and when the Wizard observes an emotional expression and interprets it as a request for help.

Each of the three games consisted of seven target words, with three possible clues for each word. A full set of clues used is available in Appendix E. As an example of a set of clues, see the illustration above in Figure 12. The target word in this case is 'good'. The picture clue shows some one giving a 'thumbs up' sign. The more difficult textual clue consists of a dictionary definition and reads 'of a favourable character or tendency: suitable or fit: commercially sound: agreeable, pleasant'. The easier clue is simply 'opposite of bad'. In some cases the words could have more than one meaning and challenge was added by having pictures and text clues that were not congruent. For example for the target word 'lies' the picture shows a figure that lies down, while the text clues are 'to make an untrue statement with intent to deceive, to be or to stay at rest in a horizontal position or be prostrate' (from dictionary definition) and 'if you were having an affair, you might tell these to your partner'.

The game is designed with changes in pace and timing, and variations in cognitive load, to induce changes of mood (Pronin and Wegner, 2006). The game is designed to trigger episodes of user frustration and satisfaction. For example, by using WOZ techniques to vary responses, it is possible to tailor presentation of clues for individuals in selected experimental conditions, by preference for type of clue (picture or text), and by degree of help or encouragement offered in messages and by rate of supply of clues for the game. Each participant has one instance of an annoying 'system interrupt' message. The mouse operation is slightly 'sticky' to operate. Time pressure is imposed by an instruction to play against the displayed clock, with a limit of 5 minutes per game. A meaningful congratulatory message is received each time a game is completed. A maximum of three games is available.

For each participant demographic and subjective data is collected (see Appendix E.).

The performance of participants is captured via a screen recorder and mouse and key logger software, so that their use of the mouse, keyboard can be measured and so that their performance can be replayed in conjunction with the video of their emotional expressions. Video capture of the participants' performances is made throughout. It is important that these features can later be viewed in conjunction with each other, so that the context in which any emotional expressions are made can be assessed. In order to test the experimental hypotheses these videos are coded for the presence of affective expressions. My study aimed to identify the affective behaviour of participants, with potential to examine events at a deep level. As my hypotheses refer to 'increase' in aspects of behaviour, I need to quantify both the number of affective events and their valence and intensity. I also want to assess the behavioural components of events to distinguish different types or variations. This initial screening approach, allows me to review my data quickly, and then spend time on in-depth analysis in the most interesting and useful parts of the data. More details are included in 4.2.6 and in Appendix D.

4.2.6. Equipment

During each trial, users were video-recorded for 10 minutes each, using a JVC GR-D70 digital video camera and JVC miniDV tape. In total I collected 60 ten-minute samples of interaction for analysis. These were later rendered into MPG2 files at 29.97 frames per second. Videos and screen capture were edited together, so that participants' behaviour could be viewed in the context of their interaction. All subsequent footage was reviewed and behaviours of interest were identified and discussed. The experiment was run on a Pentium III PC under the Windows NT operating system. A 17" Nokia 500 XA monitor in the 1024 x 768 resolution mode was used for the user display. Screen capture was recorded using CamStudio V (2009) at 29 frames per second. Mouse trax (McGinley, 2008) was used to record mouse activity and two spy logging systems PCSpy Anytime, (PCSpy, Waresight, 2009) and SC-Keylog (SC-Keylog, Soft Central, 2009) were used to log activity. These applications record user activity such as mouse movements and keystrokes and record all that happens onscreen, without the user knowing. This data can be analysed along with video recording of the user behaviour and be used in conjunction for coding observations of behaviour. This was particularly important for the analysis phase where I wanted to consider the context of observed behaviour. I used a Local Area Network (LAN) to connect machines and a double headed graphics card feeding two monitors so that the researcher received simultaneous views of their own and the participants' screens. The game was written in Java script and Visual Basic code and used a server for delivery over Active Server Pages (ASP).

4.2.7. WOZ Procedure

Before undertaking the word-ladder task, informed consent was sought with options for degree of use of data, for example for use of images in different media, and demographic data were collected (see Appendix E). The time spent by the researcher in this data collection enabled some base measures of behaviour such as observed communication style to be logged. The degree of detail given in briefings about the purposes of the study was varied according to the experimental condition as described in the factorial design. In line with ethical concerns about misleading participants, I was careful to ensure that although details were omitted in some conditions, no participants were given false information about the purposes of the study. All were told that I was looking at the usability of the system, but in only half of the cases did I add that the system might respond to their emotional expressions (see briefings in Appendix E and further discussion in 4.2.7 and 5.4).

The Wizard of Oz scenario (see Figure 13.) was achieved by a careful iterative design process. The scenario was piloted with three colleagues and fine adjustments made, for example in the positioning of where the researcher sat and exact placement of the video camera to capture the user behaviours. Recording equipment was tested and timing devices synchronised. The experiment took place in a usability laboratory where two adjoining rooms, linked by a one way mirror were used, so the researcher could observe the participant without being seen. One room was reserved for the use of participants in the experiment, containing a standard desktop personal computer (PC) with a PC mounted video camera. The other room was reserved for the researcher and contained a PC with two monitors attached to display information.

The participant in each session sat in a small study-like room with a desktop workstation. The room featured what appeared to be a large mirror but which was in fact a one-way window looking into the room.

The researchers sat behind this mirror in an adjoining room only a few feet away from the participant with a clear view of their face and upper body, just as if sitting opposite them at a table. The researchers had two monitors: one monitor with a capture of the participant’s screen, so that the researcher could see everything they did online; the other with an interface that (for the ‘system acted affective’ groups) allowed the researcher to control the timing and type of clues sent to participants.

Figure 13. WOZ scenario with researcher view on left and participant’s view on the right

Two groups were left to interact with a ‘standard application’ system that only gave clues when they used the ‘help’ button to obtain a clue. For these two groups the researchers merely observed the game play.

For the other two groups the system was meant to appear to react to their ‘emotional’ behavioural displays. For these two ‘system acted affective’ groups, WOZ techniques were used to simulate the capabilities of future emotion recognition technology. The researchers could monitor and judge the affective state of the participant in real time, via the one-way mirror, and adapt the game accordingly, so that clues and messages received by the participant were tailored to their emotional displays. The researcher could select a picture clue, a difficult textual clue or an easy textual clue. At this stage, no special rationale was considered for when to send a clue – rather the wizard responded in a natural way, when the emotional expressions or behaviours of the participant intuitively seemed to indicate that they might like another clue.

The PCs were linked by a local area network so the researcher could view the participant's 'puzzle' screen and the context of the participant's interactive behaviour at all times, as well as the researcher's own 'control' screen. All interactions were videotaped for later analysis. Logs of user keystrokes and mouse movements were collected. Screen capture was used to enable later analysis of the context in which any emotional expressions occurred, or clues were provided and to record the dependent variables operative for each participant.

Participants attempted the word-ladder game, in either 'standard' or 'acted affective' mode depending on their allocation to one of the four experimental conditions. They were given 10 minutes to interact with the system. Meanwhile the researcher monitored the users and (in the 'acted affective' condition) sent clues according to the emotional expressions perceived.

After undertaking the word-ladder task, participants were asked to rate their affective state on a nine point scale using the Self Assessment Manikin (Bradley & Lang, 1994) which uses stylized figures to illustrate emotional state. They were also asked to complete a questionnaire that rated various aspects of their interaction, including whether they believed that they had shown emotion during the experiment.

As half the participants had not been initially informed about the role of affect in the experiment, there were some ethical concerns; so all these subjects were fully debriefed at the end of their participation and apologies given for any failure to fully inform beforehand. It was made clear that they could withdraw their data from the trial, if desired. None chose to do so, although two withdrew consent for use of images.

4.2.8. General Analysis

Please see Chapter Three (3.7) for summary and Appendix E for full details of the range of pre and post experiment data requested from participants. Task performance was measured by counting the number of completed 'rungs' on the ladder. Demographic data and data from activity logs were collected. Subjective satisfaction was measured from the ratings of emotional valence (from happy to sad), given on the Self Assessment Manikin. Results were obtained from scoring standardised scales and findings from observational checklists and heuristics were recorded. All resulting quantitative data was imported into the Statistical Package for Social Sciences (SPSS) for analysis. Qualitative data was transcribed for themed analysis.

4.2.9. Coding Procedure

As context is considered critical to assessing affect, videos were edited so that both the user and their screen activity could be viewed synchronously. Videos for each participant were edited to include simultaneous displays of screen capture of the graphical user interface and interaction performance of the participant, so that coders could see what the user was doing 'in the flesh' and what they were doing 'virtually' on-line at the same time. As the video camera was situated next to the screen, the main area of focus was the face and upper body. Keyboard and mouse activities were logged by spyware and pasted into the relevant transcript. Coders were not informed of the groupings of participants and the conditions they were operating under. If a coded minute of behaviour included any delivery of clues from the Wizard then the coder could deduct that the participant was acting in the 'acted affective' condition. However coders could not be sure whether the minute selected just did not happen to have any Wizard delivered clues and in any case they did not know whether the participant was expecting affective responses or not (told affective condition).

I used the Transana (Woods and Fassnacht, 2007) qualitative video analysis tool to help organise, code and analyse the video data. I set up a Tree system with a node for each participant. Each participant had a series of videos relating to him or her. Each video had a series of transcripts attached (transcripts relate to modes e.g. audible activities, visual activities and so on). Using the Transana application, transcripts can be linked by time logs to the appropriate points in the video. Clips from the video can be selected and can have keyword families attached to them. Keywords and keyword families were developed during themed analysis.

Ideally all material should be at least double coded. To add rigour and validity to the coding process a pilot phase used double coding and a subset of data relating to six participants was coded by six coders to establish a degree of inter-coder reliability (see later details in this section and Appendix D).

Time sampling was used to reduce the coding load, and two one-minute samples were taken from each interaction and imported into Transana as two sets of 60 one minute episodes (120 in all). Two one minute sets provided quite a heavy coding load, but was just manageable. The first sample set commenced after one minute's interaction, to allow the participant time to 'settle in' to the activity, and after any influence from interaction with the research team had finished. The second sample set was of the seventh minute – towards the end of the interaction period. These were selected as representative of the overall interaction and samples were compared to entire 10 minute interactions to check they were representative. Each of the videos was viewed once, to establish trends and behaviours of interest.

I borrowed from established methods in the social sciences, such as sequential analysis (Bakeman & Quera, 1995; Bakeman & Gottman, 1997), discourse analysis (Brown & Yule, 1983; Schiffrin, Tannen & Hamilton, 2001) and phonetics (Gimson, 2001) to develop a novel, inductive, top-down coding method to assess and quantify the multimodal expressions of participants. (Please see appendices C and D for more details).

Using phonetic transcription of speech, it is usual to make transcriptions with different levels of detail, for different purposes. For example a therapist or researcher might make a quick phonemic transcription of a few phrases of interest in real time, then retrospectively use an audio recording for another pass over the data, to add more detail to any phonemes of particular interest.

The main features of phonemes (known as ‘qualities’) are:
1) if they are voiced or voiceless;
2) the place of articulation (e.g. tongue, lips, palate etc.);
3) the manner of articulation (e.g. plosive, fricative etc.).
These three qualities can then be modified or quantified, if a greater level of detail is desired, for example the manner of articulation can have added information regarding dimension – such as how long the plosive phase of a phoneme lasts (as that can alter the percieved sound).

I thought a similar approach might be useful in transcription of emotion related behaviours. I can primarily identify their main features or qualities as:
1) verbal or non-verbal;
2) the main place or mode of the emotion related behaviour (e.g. body movement, paralinguistic voice features, gaze, facial etc.);
3) the manner of emotion related behaviour (e.g. direction of body movement, intonation pattern, direction of gaze or type of facial movement.
These three qualities can then be modified or quantified, if desired, for example, speed or direction of movement, intensity of positive or negative valence, arousal, etc.).

Categories and codes were then developed on three levels (see Appendix D for more examples and details):

1. First-level coding required identification of instances of mutually exclusive and exhaustive emotional expressions or communicative affective behaviours, which were stored as clips and scored for valence and intensity. From the video footage, coders simply marked incidences that were judged to be affective related behaviour episodes involving emotional expression (which I called ‘affectemes’ in this coding system). Coders were requested to time-stamp any points where they perceived any changes in communicative signals. Observers were able to identify affective related behaviour episodes (affectemes) with ease and mark their length with distinct break points (see Appendix D and discussion in Chapter 6.4). In practice, coders often intuitively made inferences about participants’ underlying emotional state, such as

'he looks really fed up now' and/ or commented on their reasons for selecting a clip as distinctive, e.g. 'she did something with her eyes there'. These comments were collected in transcripts along with the time codes. Each point was then revisited to establish start and end points of the affecteme and each affecteme stored as a clip. I thus identified and marked the series of discrete multimodal expressions that each participant used. Samples were first generally reviewed and then meaningful episodes of affective behaviour were flagged and discussed between myself (as Wizard) and my fellow research students and supervisor. Behavioural responses of interest were categorised using intuitive names such as 'chin dump' to facilitate recognition and coding. Common behaviours such as smiling, frowning, resting chin on hand, shifting posture, grooming (e.g. adjusting hair) and blinking were identified, and counted. Each of these video clips (of affectemes) was rated for valence and arousal levels using nine-point scales, borrowed from Bradley and Lang's (1994) Self Assessment Manikin. It was important for me to quantify the valence and arousal levels of behaviours, in order to answer my research questions and test my hypotheses (from section 4.2.1).

2. Second-level coding required analysis of use of different primary modes of communication within affectemes (see Table 5). Possible modes were audible activity, whole body movements, head movements, upper face movements, lower face movements, gaze and blink patterns, and activities such as keyboard entry and screen activity. Each clip was tagged with the keyword of the primary or dominant mode used in it. In subsequent passes clips were tagged with keywords of any lesser modes. So for example although the primary mode might be angry keying in of data, the clip might include portrayal of an intense facial expression and an affect burst such as a tutting noise. I was able to look for one mode of behaviour at a time, in a systematic and thorough way, using a series of seven passes over the video footage, looking for incidences. A series of smaller modal sub-clips were made. I attached appropriate key words to each of the modal sub clips. Observers were able to identify different types of modes. Usually one affecteme clip contained several different modes, for example displaying a chin dump, a peer, an episode of rapid blinking and a sigh all within one short episode.

Transcript	Keywords – affectemes	Keyword families – examples of allaffects
1 audible activity	speech	
	whisper	
	affect burst	groan, whistle, tut, sigh, indrawn breath, snort
	extraneous noises	Foot tap, finger tap, door bang
	data entry	keyboard, mouse click
2 whole body movement	shifts	large shift, small shifts, lean back, postural twirl
	tension	hunched shoulders
	grooming	tuck hair, shunt glasses, bite finger/thumb, scratch
	chin dump	L hand, R hand, both hands, mouth covered, fisted, fingers spread
3 head movements	gesture	nod, shake
	peer	distant, close
	aspect shifts	tilt, turn, L, R, up, down
	chin move	tuck, thrust
4 upper face	brow-raise	bilateral, unilateral L, unilateral R
	frown	slight, deep
	nose	wrinkle, flare
5 lower face	smile	PanAm, zygomatic
	mouthing words	
	fidgets	compress, pursing, rinsing, jaw grind
6 gaze and blinks	eye shifts	flashbulb eyes, narrowed, closed
	screen attention shifts	on-screen, off-screen, scanning, L, R, up, down
	blinks	
7 keyboard, mouse and on-screen activity	data entry	Keystrokes, mouse clicks
	on-screen activity	Picture clue, short text clue

Table 5. Second-level transcripts and keywords used.

3. Third-level analysis required categorisation of incidences within these second-level ‘families’, for example the family labelled as ‘grooming activities’ included biting fingers, scratching head, or adjusting glasses or hair.

The first analysis level consisted of discrete episodes, but coding of instances within levels two and three could co-occur. Coding enabled me to use a themed analysis technique (following Bakeman & Gottman, 1998) to develop sets of keywords that could be attached to time coded behavioural segments. A coding protocol was designed to log the number of affective related behaviour episodes (or ‘affectemes’ as I called them in this coding system). In Transana the keyword families of modes were set up for attachment to each transcript. The transcripts and keyword families were set up for the three levels of operation, relating to the qualities of the affectemes, such as their mode and manner and modifers of those (see Appendix D.). Once coded, data was imported into the Statistical Package for Social Sciences (SPSS) for analysis. The experiment’s main research hypotheses were tested using analysis of variance (ANOVA).

4.2.10. Reliability of the Method

My research question requires identification of emotion-related episodes of behaviour (affectemes) and rating of the valence and arousal levels displayed by participants during them. I need to check that different raters identify episodes and ratings consistently. I followed three phases to establish agreement. First, I examined a small sample of data in some depth to establish initial criteria and identify affectemes. Secondly, I ran a trial study to code data for valence and arousal, comparing results of pairs of coders, to inform what additional training coders would require and what refining criteria and procedures needed, as calculations for each coder pair can be a valuable diagnostic tool during coder training (Lombard, Snyder-Duch & Bracken, 2002). Thirdly, I ran a pilot-coding exercise to establish that reliability was satisfactory across multiple coders (see following section and paper in appendix D for more details).

Reliability Agreement - First stage

For the first stage a one minute pilot sample was rated by three coders (including the researcher) and differences were noted and discussed. It proved intuitive to identify segments where a discrete message or episode of affect was perceived, although sometimes identification of onset and offset times varied by a few seconds, as identified by different coders. In some instances one coder would initially identify fewer segments as occurring than another, although when segments were discussed, then there was agreement that two segments could be construed as parts of a larger meaningful segment. I achieved percentage agreement levels of over 90%, and Krippendorff's alpha .9740 for agreement on identifying affective segments. Episodes of affective behaviour were discussed and named. Using Transana qualitative analysis software video clip collections were built, which allowed clips to be viewed in sequence, for example all 'chin dumps' can be reviewed, to check that they are all instances of the same behaviours.

Once segments were identified, than each segment was then rated for valence and intensity. These ratings were transcribed for each second of activity. Confusion matrices were quickly drawn to provide quick visual comparisons of results. This showed that although ratings were not exactly the same, ratings were usually in the same quadrant.

The Self Assessment Manikin proved very helpful and intuitive to use as an assessment tool. Using this 9-point scale, ratings were often similar, falling within the same range, but not always identical, as shown in Table 6 and 7. It is clear from Table 7 that coders tend to perceive similar main boundaries between affectemes, for example all perceive a change in intensity and valence between 40 and 45 seconds, but they vary in their view of exact second of onset and offset of the affecteme and in their exact rating of the intensity and valence. Also some coders seem to perceive some affectemes as composed of smaller sub units, with more changes in intensity and valence when compared to others.

Second Number	91	92	93	94	95	96	97	98	99	100
Rating given by Rater 1	7	7	7	7	7	7	4	4	4	4
Rating given by Rater 2	7	7	7	7	7	7	4	4	4	4
Rating given by Rater 3	7	7	7	7	6	6	6	6	6	6
Rating given by Rater 4	6	7	7	7	5	5	5	5	5	5

Table 6. Sample 1 of coding of data

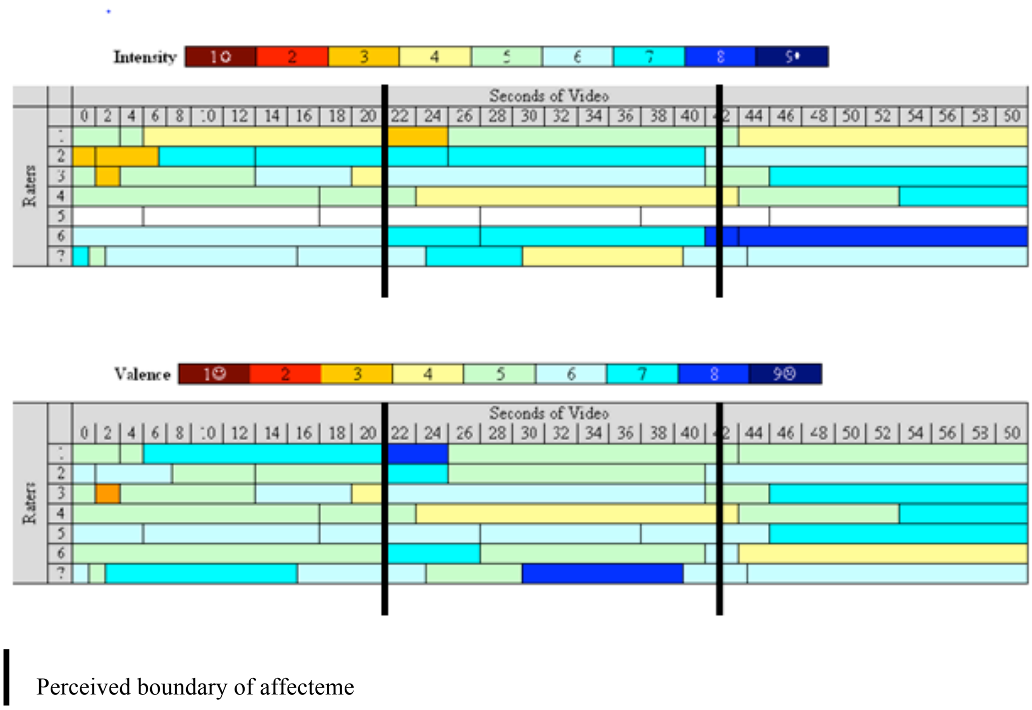


Table 7. Sample 2 of coding of data, showing boundary similarities

In order to assess the reliability of my coding of content analysis, I need to assess more than correlation or analysis of variance values. I need to assess the inter-coder agreement that is the extent to which the different judges tend to assign exactly the same rating to each object. Lombard, Snyder-Duch et al (2002) reviewed literature regarding inter-coder reliability and found poor use and reporting of methods. They developed useful guidelines for appropriate procedures, followed, here, which recommend reporting actual inter-coder agreement for variables on coded values which should be the basis for assessment used (even if, as with my data, similar rather than identical values 'count').

I selected several indices to measure agreement. First I considered agreement between pairs of raters using percentage agreements, and Cohen’s Kappa to account for agreement expected by chance between pairs of raters. Despites some drawbacks, this is regarded as the “measure of choice” for rating behaviour by Bakeman (2000). I then used Krippendorff's alpha to consider coefficients between multiple raters.

Reliability Agreement - Second stage

For the second stage, a representative sample of my entire data was selected to be tested for pilot reliability testing consisting of 6 one minute samples of data (10% of total data). These samples are a subset of my main data, selected to represent the main conditions of my experiment and to represent both male and female participants. I used 5 coders to code these samples. Each coder received a 10-minute training session, reviewing a one-minute sample with the researcher and discussing the coding criteria. At this stage I compared results between pairs of coders for their ratings of each 60-second sample, giving six comparisons per pair of coders per variable.

I used Bakeman and Quera’s (1995) Generalised Sequential Querier, ‘GSEQ’ software to calculate percentage agreements and Cohen’s Kappa between pairs of coders and I used the Statisitcal package for social sciences software, (SPSS) with a macro for Krippendorff's alpha (Hayes, 2005) with which computes alpha for any level of measurement and any number of judges.

Lombard, Snyder Duch & Bracken, (2005) state that "there is general consensus in the literature that indices that do not account for agreement that could be expected to occur by chance (such as percent agreement…..) are too liberal (i.e., overestimate true agreement), while those that do can be too conservative." Cohen's kappa and Krippendorff's alpha do not allow any credit for ratings which are close, but not exactly the same, although in this study, if one rater put valence at 7 and one at 8 that might be considered close enough. Although my percentage agreements were good, when the more stringent tests were used, the reliability scores were lower.

For valence ratings, my percentage agreements ranged from 83.33 to 100% with a mean of 92.2%. Kappa agreements ranged from 0.11 to 1.0 with a mean of 0.58. Krippendorff's alpha was 0.2301.

For arousal ratings my percentage agreements ranged from 83.33 to 97.78% with a mean of 89.8%. Kappa agreements ranged from 0.13 to 0.86 with a mean of 0.45. Krippendorff's alpha was 0.0232.

These results did not give good enough agreement levels and did not meet my agreement criteria. I hoped to achieve percentage agreements of 75% or above, as this is a liberal agreement index (Lombard, Snyder-Duch et al, 2005). Using the more stringent and conservative Cohen's Kappa or Krippendorff's alpha I would be satisfied with relatively low coefficients of .70 or greater as this is an exploratory study (Lombard, Snyder-Duch et al, 2002).

I examined the results for pairs of coders in detail to try to understand why I had not met my criteria. It was interesting to note that there was higher agreement for valence than for arousal ratings, and that there was higher agreement between same sex pairs of raters, particularly when female and particularly when rating the opposite sex.

Reliability Agreement - Third Stage

In view of the stage two results I decided to test for reliability using four female raters and a longer training session lasting one hour, during which six one minute samples (3 male and 3 female) were examined in detail. Each rated the 6 samples used previously, following the same procedure as before. Ratings were closer than in the previous pilot. Whereas ratings had sometimes differed by up to 4 points on the rating scale in the pilot ratings, they were now within 2 points in all cases.

I used 360 seconds of data from the 6 one minute samples for valence, to compute Krippendorff's alpha and repeated the exercise using arousal data, so that there was one comparison for each set of coders for each variable. Krippendorff's alpha met my criteria for agreement with valence rating agreement of 0.7836 and arousal rating agreement of 0,7165.

This difficulty in establishing inter-coder reliability has important implications. If we hope to design and develop systems with human-like capability for emotion recognition, then we must first understand the variation in human performance. And as affective recognition systems are still in their infancy and are proving so complex to develop, it is likely that we will need to use human coders to at least validate performance of systems for some time to come and we will need to know that our human coders are reliable. Different gender dyad combinations may affect outcomes.

I carried out retrospective walkthroughs to see how my participants rated their own behaviour in comparison to the coders (described in 5.2).

4.3. Summary

I used a WOZ approach to successfully simulate the capabilities of future affective systems. The Wizard of Oz technique used here appeared to work successfully to simulate the capabilities of future affective systems and could be used in the future to explore further questions within this domain.

By adopting a user centred approach, using empirical methods, and experimental and factorial design I was able to design an experiment with 2x2 factorial design that allowed me to compare performance and behaviour of groups of participants. The first factor was to compare groups who were interacting with a system that varied in its response to them so that it did or did not appear to respond to their emotional expressions. The second factor was that only half of each group were told to expect an affective system, resulting in four groups for comparison. This variation in recognition of emotion plus or minus the intervention resulted in a rich and interesting data set. A systematic approach with structured qualitative methods (influenced by the social sciences) and use of trained and experienced observers make observational analysis as objective as possible. I developed a novel coding system using a macro-analysis approach with different levels of coding. Detailed analysis of the observational data has the potential to further inform future design work.

The development of hypotheses and then the design and implementation of an affective game, the implementation of the experiment and testing of groups under different conditions, and validation and analysis of data, proved challenging and required consideration of many complex aspects on both theoretical and practical levels. There is a real danger of bias creeping into the execution of the experiment, so it is important that the use of experimental design allows for control and hunts for bias in a number of ways. For example, I control the make up of groups to ensure balance between groups. Also the information given to all four groups of participants is standardised in that it involves convincing them of their surveillance and the presence of the video camera is overt. As they all believe that they are under observation, the Hawthorne effect will affect all conditions equally.

Chapter five describes the results of this experiment this study.

5. Affective Experiment Results

5.1. Introduction

This chapter describes the analysis and results of the main experimental work resulting from the experimental design described in the previous chapter. In section 5.2 the results from initial coding procedures are described, showing the kinds of interactions, rich in the use of emotional expressions, that took place during the experiment.

Then in 5.3 the main results form testing my experimental hypotheses are presented, that show effects on task performance, subjective satisfaction and on ratings of valence and arousal changes in the emotion related behaviours that the participants exhibited or described. The results are explained with illustrative examples and report details of statistically significant gains in task performance, and subjective satisfaction that suggest there are real benefits to participants who use the affective condition. It describes my findings relating to valence and arousal levels and to observed behaviours. As well as these quantitative results, findings from the qualitative analysis of data collected from pre and post experiment questionnaires are described.

In section 5.4 I discuss some of the implications of the results. The experimental design allows drawing of powerful conclusions around the cause and effect of the benefits found in terms of task performance and subjective satisfaction.

In Section 5.5 I go on to draw some conclusions from the study and identify interesting issues and areas for further study and post hoc work.

5.2. Initial Results from Coding Procedure

First-level coding – sample results. Initial screening showed that it was not always easy to see my participants' emotional expressions because they sometimes moved off camera or because of masking by wearing glasses, baseball hats, 'hoodies', or interference from extraneous activities such as chewing gum. Despite these real world challenges, I identified a wide range of behaviours by individuals (see examples for one participant in figure 14). Occasional smiles and frowns were obvious indicators of affect, which along with occasional grimaces appeared as occasional floods of emotion crossing the face. Episodes of shifting posture – either in discomfort or in readiness for an item were also noticeable. A striking number of participants spent significant time resting their chins in their hands while they thought how to proceed. There was noticeable, habitual ‘grooming’ behaviour by some participants, such as tucking their hair behind their ears or adjusting their glasses. Some participants were quite vocal (although there was no speech interface) and talked aloud to themselves, whispered clues aloud and many made non-speech noises or affect bursts such as sighs, ‘tutting’ noises, whistling, etc. Episodes of rapid blinking were also noted. Silent mouthing of clues was common, and mouth tension was displayed by fidgeting movements such as jaw grinding, lip pursing, lip licking, lip compression or mouth-rinsing actions. Some common affecteme patterns were recognised and intuitively named. For example, the ‘anxious peer’, where participants peered at something on-screen for several moments or, the ‘getting involved postural twirl’ as participants shifted around and then returned to a preferred position (often a turn-taking cue in human-human communication).

Figure 14. Array of emotional expressions of one participant

It was noticeable that coders’ judgements of affect altered over time and depending on context. For example, when participants sat still without any emotional expressions while reading a new clue – this seemed comfortable and positive for a short time. However, if they could not then guess the clue and proceed with the game, the same expression started to look more negative. Gaze was also significant for judgements of engagement and affect, and it was noticed that some individuals seemed to blink excessively at times during interaction.

From the second set of interaction samples (the seventh minute) 401 affectemes from 60 different participants (mean 6.68) were identified. The mean length of an affecteme was 9.02 seconds with a mode of 3.7 seconds and a range of 0.1 – 48.3 seconds.

Each affecteme was rated for valence and intensity using an adapted version of the Self Assessment Manikin scale. This found a mean intensity rating of 4.89 with a mode of 5 and a range of 2 – 8, where 1= high arousal and 9 = low arousal. The mean valence rating was 5.74 with a mode of 6 and a range of 3 – 8, where 1= positive and 9 = negative.

Second-level coding – sample results. For example, during the first set of one minute interaction samples (taken after one minute) the following were noted:

- 53 episodes of 'chin dumps' where participants rested their head in their hands, (mean 0.88 per participant). The total duration was 581 seconds (mean duration 9.68 seconds per participant);
- 287 episodes of postural shifts (mean 4.78 per participant). The total duration of these was 517 seconds (mean 8.62 seconds per participant);
- 29 episodes of grooming in the form of adjusting hair, clothing or spectacles, etc. (mean 0.48 per participant) with total duration of 65 seconds (mean duration 1.08 seconds per participant);
- 69 vocalic instances, such as brief sighs, groans or laughs (mean 1.15 per participant);
- 803 instances of blinking (mean 13.38 per participant).

There were individual differences between participants, so that not all displayed all possible behaviour components:

- all 60 participants used postural shifts (duration range 1 – 41 seconds, mean 8.62);
- 32 participants used chin dumps (duration range 1 – 47 seconds, mean 18.16);
- 31 participants used vocal sounds, mainly affect bursts (range 1 – 6 instances, mean 2.23);
- 12 participants used grooming behaviours (range 1 – 5 instances, mean 2.42).

Third-level coding. For further analysis at the third level of coding, when required, my top-down approach allowed me to code variations on an affecteme and episodes could be further categorised within the keyword family, so that the type of smile, or exact nature of a chin dump could be transcribed by adding appropriate keywords from that family. For example, a chin dump might involve: one or both hands; contact with the cheek or the chin; contact of whole hand palm or fingertips; covering of mouth, mouth and nose or neither. For the purposes of my analysis these could all be classed as the same behaviour, although fine details might vary, as shown in figure 15.

In phonetics, variations on a phoneme often occur. For example the quality of the 'l' sound at the beginning and end of the word 'little' is quite different (known as dark and light 'l'). These variations within phonemes are called allophones. In parallel with this, I suggest that variations on an affecteme could be called allaffects.

Figure 15. Variations on a chin dump. Allaffects

Coding validation and retrospective comparison

I carried out considerable work to validate my coding (see discussion section 6.4) and I carried out retrospective walkthroughs to see how my participants rated their behaviour in comparison to the objective raters. I found good agreement between participants and raters about onset and offset of events. Rating agreements using the Likert scales were often just one or two points apart, with valence more easily agreed than arousal. There was closer agreement about ratings at extremes, for example very positive or negative, but less agreement about more neutral states. There was also some interesting discussion about appraisal of the participants' behaviour. Different uses of affective terminology for emotional states sharing the same general valence was striking. For example, on one event where the rater had described the participant as 'anxious', the participant described himself as 'confused' and 'angry'.

5.3. Results

When the video footage obtained from my study was first reviewed, it was clear that participants' behaviour was rich in communicative, affective behaviour episodes that were perceived by the researchers as indicating episodes of disaffection, tension or enjoyment. It was felt that some sort of complex message was being given at all times. Although events did not always contain overtly affective behaviour, and rarely showed prototypical facial expression of basic emotions, behaviour episodes could still be identified as relating to affective states. In turn the affective states could be assumed to be induced or influenced by the interaction experience.

5.3.1. Task Performance

Hypothesis 1. If the system acts affectively, there will be improved performance by the participant.

The task performance was measured by seeing how many 'rungs' of the first word ladder game each participant completed. Overall, the mean number of rungs completed was 6.48, SD 1.43. For the different conditions, the means and standard deviations were as shown in the following table.

	Application appeared to react to emotional expressions	Standard application
Participants told application affective	Group 1 Mean 7.0 SD 0.0	Group 2 Mean 6.47 SD 1.24
Participants not told	Group 3 Mean 6.93 SD 0.26	Group 4 Mean 5.53 SD 2.36

Table 8. 2 x 2 factorial design and number of rungs completed

Analysis of variance (ANOVA) was carried out and showed a main effect of system affective response on task performance ($F (1,56)=7.82$, $p<0.01$). Participants were able, on average, to complete significantly more rungs of the puzzle with the affective intervention, when clues were supplied not only in response to the participant pressing the 'help' button, but also by the Wizard responding to their emotional related behaviours.

5.3.2. Subjective Satisfaction

Hypothesis 2. If the system acts affectively there will be improved satisfaction on the part of the participant.

The participants' satisfaction was measured by seeing how they rated their affective state immediately after the interaction, using the valence scale of the Self Assessment Manikin (where 1 is very positive and 9 is very negative). Overall, the mean rating was 4.02, SD 2.10. For the different conditions the means and standard deviations were as shown in the following table.

	Application appeared to react to emotional expressions	Standard application
Participants told application affective	Group 1 Mean 3.47 SD 1.80	Group 2 Mean 4.73 SD 1.83
Participants not told	Group 3 Mean 2.93 SD 1.79 (most positive group)	Group 4 Mean 4.93 SD 2.40 (most negative group)

Table 9. 2 x 2 factorial design and subjective satisfaction

There was a main effect of affective response on user subjective ratings of their affective state after the interaction (F (1,56)=10.25, p<0.005). Participants reported themselves as significantly happier, on average, after interaction with the system that responded to their emotional expression.

5.3.3. Valence

Hypothesis 3. If the system acts affectively, then participants' expressions of emotion will increase in valence.

The participants' valence was measured, following the interaction, by the researchers rating the affectemes during the one-minute sample video clips, using a scale based on the Self Assessment Manikin (where 1 is very positive and 9 is very negative). Overall, the mean rating was 5.47, SD 0.72. For the different conditions the means and standard deviations were as shown in the following table.

	Application appeared to react to emotional expressions	**Standard application**
Participants told application affective	Group 1 Mean 4.73 SD 0.38 (the most positive group)	Group 2 Mean 5.80 SD 0.75 (the most negative group)
Participants not told	Group 3 Mean 5.67 SD 0.54	Group 4 Mean 5.69 SD 0.62

Table 10. 2 x 2 factorial design and valence

I found that when using an apparently affective system, users' affectemes are rated as having significantly more positive valence. ANOVA showed a main effect of system affective response on ratings of valence of participants' emotional expressions. F(1,56)=12.63, p<0.01). Participants showed, on average, use of more positively rated emotional expressions with the affective intervention. Participants were more likely to show positively rated emotional expressions when they had also been told that the system was affective. There was a significant interaction effect with whether the system had provided an affective response (F(1,56)=12.63, p<0.01). The most positively rated emotional expressions were from those participants who had been told that the system might respond to their emotional expression, and where the system did in fact act affectively (interaction effect plot in figure 16).

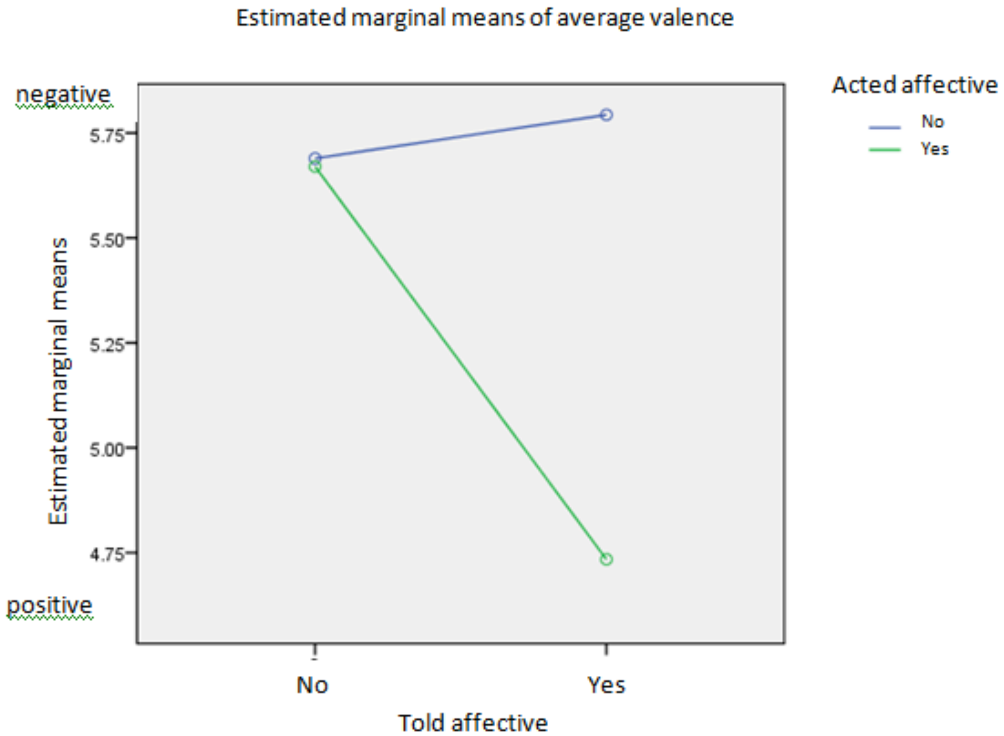


Figure. 16. Interaction plot for ratings of positive valence

5.3.4. Arousal

Hypothesis 4. If the system acts affectively, then participants' expressions of emotion will increase in intensity.

The participants' arousal was measured, following the interaction, by rating their affectemes during the one-minute sample video clips, using a scale based on the Self Assessment Manikin, (where 1 is very high arousal and 9 is very low arousal). Overall, the mean rating was 4.38, SD 0.74. For the different conditions the means and standard deviations were as shown in Table 11. I found that when told the system they will be using is an affective system, users' affectemes are rated as having significantly more intense arousal (although only just significant). ANOVA showed a main effect of system affective response on ratings of valence of user's emotional expressions. $F(1,56)=4.74$, $p<0.05$). Participants showed, on average, use of more intensely rated emotional expressions with the told affective intervention. Participants were more likely to show emotional expressions rated as having higher arousal when they had been told that the system was affective.

	Application appeared to react to emotional expressions	**Standard application**
Participants told application affective	Group 1 Mean 4.12 SD 0.42 (highest arousal)	Group 2 Mean 4.22 SD 0.79
Participants not told	Group 3 Mean 4.52 SD 0.89	Group 4 Mean 4.64 SD 0.73 (lowest arousal)

Table 11. 2 x 2 factorial design and arousal

5.3.5. Behaviours

Hypothesis 5. If the system acts affectively, then emotion related behaviours will change.

In order to test this hypothesis I needed to systematically cover each independent variable. The required evidence was collected from the systematic observations and from the post experiment questionnaire data:

- The number of observed behaviours in each minute was counted to see if number of observed overall behaviours increased in any conditions;
- Counts of observed behaviours between groups and from first and second minute samples were compared to see the effect of different conditions and whether there was any progressive change in behaviour over time.
- Examples of behaviour types were qualitatively and quantitatively analysed to see if number or quality of behaviours altered in in different conditions;
- Participants self reports of degree of emotion they felt they expressed (taken from post experiment questionnaire, Appendix E) were analysed to see if participants themselves felt that their display of emotion related behaviours changed according to the experimental condition;
- Numbers of clues supplied in different conditions were counted and compared.

Overall, the mean number of affectemes per participant for the two minutes coded was 6.48 with SD 2.19. For the different conditions the means and standard deviations were as shown in the following table (Table 12.).

	Application appeared to react to emotional expressions	**Standard application**
Participants told application affective	Group 1 Mean 7.30 SD 2.43	Group 2 Mean 6.37 SD 2.21
Participants not told	Group 3 Mean 5.83 SD 1.84	Group 4 Mean 6.43 SD 2.19

Table 12. 2 x 2 factorial design and number of affectemes

Comparisons between the first minute and second minute were as follows. The group who were not told that the system might react to them, but found that it did appear to react to them, increase the number of behaviours during the second minute sample compared to the first, as shown in the following table.

	Application appeared to react to emotional expressions		**Standard application**	
	1st sample	2nd sample	1st sample	2nd sample
Participants told application affective	Group 1 Mean 7.20 SD3.53	Group 1 Mean 7.40 SD 3.11	Group 2 Mean 5.87 SD 3.07	Group 2 Mean 5.80 SD 1.66
Participants not told	Group 3 Mean 5.73 SD 2.63	Group 3 Mean 7.00 SD 2.42	Group 4 Mean 6.33 SD 3.11	Group 4 Mean 6.53 SD 2.32

Table 13. 2 x 2 factorial design and comparison of number of affectemes in 2 samples

Individual participants show an immensely wide array of emotional expressions during affective HCI, ranging from subtle to intense and using many modalities, in all experimental conditions during the course of a game. Although not all participants used all behaviours, some modes and expressions were very common. Interestingly, since it was a problem solving game, the stereotypical pose of a thinker, with chin resting on hand, was frequently used. I found that sometimes there were variations on an affecteme that did not alter meaning significantly. For example whilst resting chin on hand they might use the right or left hand. The hand might be positioned laterally or centrally. The fingers might be curled or stretched. None of these variations significantly altered the perceived emotional message. Although it would be possible to analyse behaviours in greater detail, it was not required for the purpose of my study.

Facial expressions, such as smiles and frowns were used for various durations. Postural shifts – both large and small were noticeable, with frequent attempts to peer closely at the screen as if to clarify the task. Although the problem solving game did not use a speech interface, some users did speak aloud or whisper in order to read clues or to reason with themselves. They also used a wide range of affect bursts such as sighs, laughs and whistles. Mouth tension was often evident with jaw grinding or lip pursing and rinsing motions carried out. Grooming activities, such as tucking hair behind ears or shunting spectacles was also frequent.

Overall, the mean score of participants for 'felt they had showed their emotions' was 0.72 with SD 0.45. For the different conditions the means and standard deviations were as shown in the following table.

	Application appeared to react to emotional expressions	**Standard application**
Participants told application affective	Group 1 Mean 0.73 SD 0.46	Group 2 Mean 0.93 SD 0.26
Participants not told	Group 3 Mean 0.73 SD 0.46	Group 4 Mean 0.47 SD 0.52

Table 14. 2 x 2 factorial design and extent participants believed they showed their emotions

There was a main effect of being told that the system was affective on participants' self reporting that they had displayed their emotions ($F(1,56)=4.34$, $p<0.05$). Participants were more likely to say they showed their emotions when they had been told that the system was affective. There was also a significant interaction effect with whether the system had provided an affective response ($F(1,56)=4.34$, $p<0.05$). The highest self-report of showing emotion was from those told that the system might respond to their emotional expression, but where the system had not in fact been affective. This interaction effect is shown in the interaction plot in figure 17.

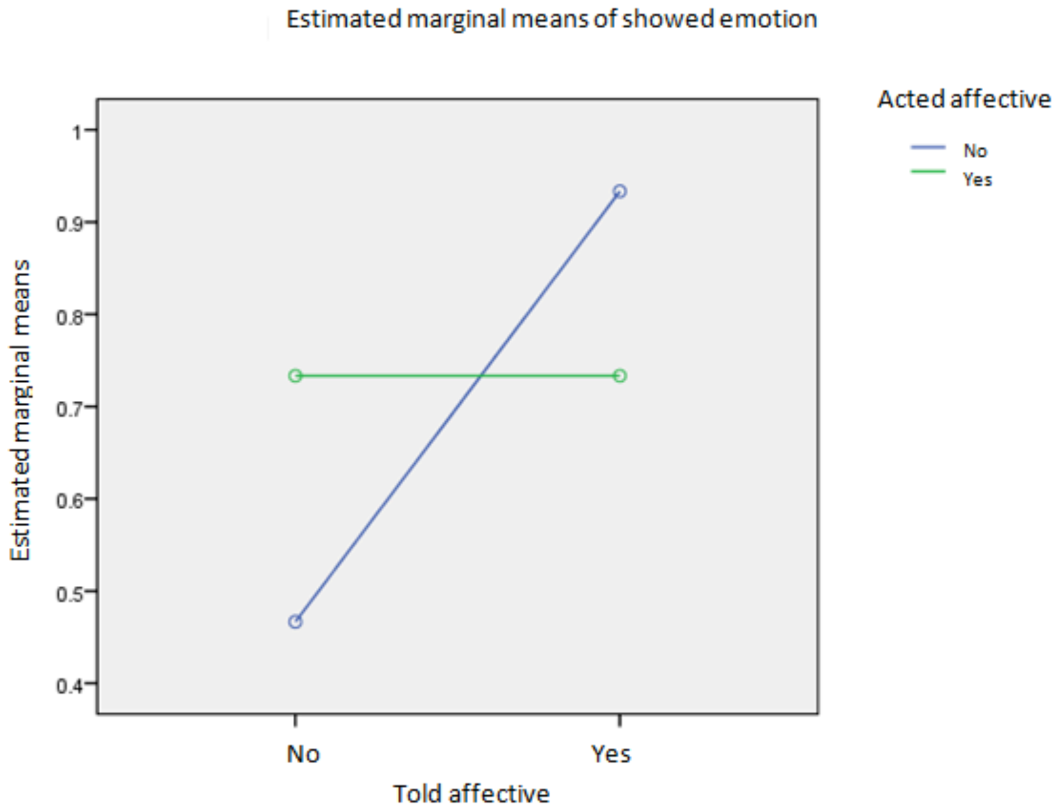


Figure 17. Interaction plot for self reported emotion display

Of the macro-level affective behaviours that have been coded, only blink rate was found to vary with experimental condition. Overall, the mean self-rating of the extent to which participants blinked for the two minutes coded was 13.38 with SD 9.28. For the different conditions the means and standard deviations were as shown in the following table (Table 15.).

	Application appeared to react to emotional expressions	**Standard application**
Participants told application affective	Group 1 Mean 17.13 SD 9.09	Group 2 Mean 14.67 SD 10.99
Participants not told	Group 3 Mean 10.33 SD 9.08	Group 4 Mean 11.40 SD 6.82

Table 15. 2 x 2 factorial design and blink rate

There was thus a main effect of being told the system was affective on blink rate ($F(1,56) = 4.57$, $p<0.05$). Blink rates were significantly higher when participants were told the system might respond to emotional expressions.

Overall, the mean number of clues (see section 4.2.4) requested was 2.05 with SD 1.25. For the different conditions the means and standard deviations were as shown in the following table.

	Application appeared to react to emotional expressions	**Standard application**
Participants told application affective	Group 1 Mean 1.80 SD 1.08	Group 2 Mean 1.53 SD 1.41
Participants not told	Group 3 Mean 2.40 SD 1.24	Group 4 Mean 2.47 SD 1.12

Table 16. 2 x 2 factorial design and clues given

There was a main effect of being told that the system was affective on participants' number of requests for clues ($F(1,56)=5.91$, $p<0.05$). If they were told the system was affective, they were less likely to ask for clues.

5.3.6. Qualitative data from questionnaires

Some qualitative data (see appendix G) was collected, after the game play, via the post experiment questionnaire (see Appendix E), regarding emotional experiences and attitudes. The post experiment questions asked participants about their emotional experiences (see Appendix E). A themed analysis was carried out to establish themes and concepts expressed by the participants.

Questions included:

1. During the 'word ladder' game, did you feel any emotions? Yes / No

If so what did you feel?

2. If you felt any emotions, do you think you showed them in any way? Yes / No

If so how did you show them?

3. Did the word ladder game seem to adapt itself to you in any way? Yes / No

If so how did it adapt?

4. Do you believe that the computer recognised any of your emotions? Yes / No

If so what emotions did it recognise and how do you think it recognised them?

5. Do you believe that the computer responded to any of your emotions? Yes / No

If so how did it respond? …………………………

Q1. Feeling emotion When asked if they had felt any emotions, all four groups made comments about the mixed emotions they had felt, reporting far more negative (72) than positive (30) experiences, for example:

> "I feel nervous before the game and stupid when my mind went blank"
>
> "Enjoyment when I got it right, frustration and embarrassment when I couldn't think of the right answer"
>
> "A little frustration when I got stuck, followed by amusement at myself because it is only a game"
>
> "Slightly frustrated and anxious but happy when I got the answer right."

Q2. Showing emotion. 43 of the 60 participants believed that they had shown emotions in some way. Those in the 'not told affective' groups gave fewer responses and those in the control group (not told affective and standard application) gave least responses of all. Interestingly two participants from groups told the system might react to their emotions, set out to intentionally refrain from showing emotion, although in fact my video analysis did detect affectemes, saying:

> "Yes, felt them - but I never showed up any expression."
>
> "It didn't get anything out of me - I kept a blank face - so it couldn't tell if I was feeling any emotions anyway. I controlled my face so it wouldn't be able to respond to me."

One participant, told the system might react to displays of emotion, claimed:

> "I gave up trying, because it was annoying."

The participants described showing emotions via:

- facial expressions (39 mentions), including laughing or smiling, frowning, biting lip, gaze direction, etc.;
- body movements (14 mentions) including fidgeting, bouncing on the chair, moving to and fro, etc.;

- hand movements (13 mentions) including gestures, tapping fingers, hitting keys harder, moving the mouse, touching their face or head, etc.;
- audible behaviours (12 mentions) including laughing, muttering or mumbling, swearing, reading clues aloud, etc.

Q3. Game adaptation. Asked if the game had adapted to them, in any way, around one third (18 out of 60) participants agreed, with the comments they made split between those who thought adaptation was made by the game (10 comments), and those who thought that they themselves had adapted their behaviour (9 comments). A number (8) made comments to show they ascribed the simplicity of the game design as affording their adaptation. The control group, not expecting adaptation and using a standard non affective system gave least responses (response by only one person). The groups told to expect an affective system made most comments and those for whom it did react with apparent affect made most of all.

Q4. Emotion recognition.

Only 17 out of 43 participants believed the computer had recognised their emotions, with the majority of these in the groups told to expect affective reaction. 10 participants made comments on how the computer had detected a behavioural change in them, all naming behaviours with negative connotation, using words such as "confused", "puzzled", "difficult", "frustration", "nervous", "idle", "struggle", "couldn't get it", for example:

> "I was confused in between so may be that one."

11 made comments on possible detection mechanisms based on the way they played the game that had been interpreted as revealing emotion, for example:

> "I noticed that the time calculator stopped whenever I asked for help. That might be a kind of recognition of my emotions."
>
> "By the fourth 'rung' down I'd get slightly more confused."
>
> "Frustration - by the amount of time I took to respond to the clue."

One assumed the video camera was being used to capture emotion in some way and one pointed to a perspex button on the computer tower and asked, "Is that the emotion sensor?" The 'not told affective' groups gave the least responses and the 'not told affective and standard application' control group least of all.

Q5. Response to Emotion: Less than one third (13) believed that the computer responded to their emotional state by adapting the game to help them. The group told to expect an affective response and for whom there was an apparently affective response, made by far the most comments. 6 respondents suggested emotions that had been detected, including “troubled”, “confused”, “feeling stuck”, “not knowing”, and rather mysteriously and all encompassingly:

> “Some times they (computers) can sense what or how you are feeling.”

Participants mentioned types of system responses they had noticed, including the supply of help messages or hints; questions; or answers that varied in tempo or quality, including one who one thought the system made questions harder! Some thought these changes related to specific points or levels in the game, such as:

> “On task two it gave an answer much quicker when I pressed help.”
>
> “After a few times of pressing the help button it started showing pictures next to the word ladder to help me.”

Development of emotion recognition systems: My participants were largely in favour of the development of affective computers that could recognise emotions (as shown in Table 17). None said ‘never’, only 4 were against it, 13 were neutral and 39 out of the 60 were broadly in favour, with slightly more of these feeling mildly, rather strongly in favour.

Some saw the rise of emotion recognition as inevitable, making remarks such as:

> “Well I think this is a natural step of technology since computers and machines are becoming part of our lives.”
>
> “It is likely to happen because the tech is improving every day.”
>
> “I see it as logical progression.”

50 positive comments were made, including:

> “It's an interesting way to work with computers.”
>
> “Cool.”
>
> “Sounds useful.”

> "I believe we would be very pleased that computer systems and other electronic devices are able to interact with us in a much more interesting sense."
>
> "I won't think computer game is just a game, I will feel like I'm playing wiv my friend."
>
> "It can only help people therefore it's a good idea and will save time and money hopefully."

But a lot of comments had caveats, for example:

> "Brilliant but they'll never be human."
>
> "It would be desirable in some circumstances, but not others."
>
> "Would be clever, but intrusive way of effecting someone's emotions."
>
> "May be good, if used for learning" (relating to type of application).

And a number of negative comments were made including:

> "It should not happen in the usual use of a computer because it divert the concentration of the user & feel frustrated."
>
> "I think it is pretty scary that computers are becoming more 'human' and can be adapted to recognise human emotion and error."

Finally, a number of comments were made about uncertainty, personalisation and control of systems, e.g.: "Honestly, I'm not sure."

> "If people recognise that adverts are mood targeting, they can still control what they buy."
>
> "It will cater for each user individually."

	1 Feel strongly - they should never be developed	2 Feel mildly - on the whole best if not developed	3 Don't care if developed or not	4 Feel mildly - best if developed	5 Feel strongly - they should be developed	No data
Group 1 Told and Reactive	0	1	3	5	4	2
Group 2 Told and Standard	0	1	3	6	5	0
Group 3 Not Told and Reactive	0	0	4	7	4	0
Group 4 Not Told and Standard	0	2	3	4	4	1
OVERALL	0	4	13	22	17	3

Table 17. Views on development of emotion recognition systems.

Overall The group expecting the system to respond, but for whom it did not, made the most negative remarks, including “irritated” and “annoyed”. The group for whom the system acted in the expected way (told it would be affective) contained six individuals who felt they had not felt any emotions during play.

5.4. Discussion

The results described here demonstrate that significant improvements in both task performance and users’ subjective feelings of satisfaction can be achieved as a result of adapting an application on the basis of affect recognition. Participants performed significantly better and reported that they felt more positive when using the adaptive affective recognition version of the application. This confirms suggestions that affective computing has the potential to provide positive outcomes that may improve human-computer interaction. Previously empirical support for this claim was limited due to the relative immaturity of the technology to enable emotion recognition. In this experiment I was able to simulate (using a Wizard of Oz set up) the proposed capability of future systems to recognise affect from observed user behaviour. The results are notable as they support the importance of ongoing work to enable emotion recognition at the user interface.

The participants who used a system that appeared to act affectively reported themselves as feeling the experience was significantly more positive, whether or not they knew it might act affectively. However the participants did experience significantly higher levels of arousal when they were expecting the system to act affectively, whether or not it did. These groups also achieved higher scores on the negative aspects of the affect susceptibility measure and had significantly higher blink rates. It seems that those anticipating using a novel, affective system were in a higher state of arousal, and possibly felt some added stress or felt the need to be more alert. It is not possible to say from this experiment whether the high arousal was due to the affective nature of the system or the novelty of it. Further experiments could help to establish this, either by participants using an affective system repeatedly or by comparing use of an affective system with another equally novel but non-affective system.

The research also considered how users' emotional expressions might vary when they use an affective system. It has been suggested that, in the short term at least, emotion recognition technology will need users to exaggerate or deliberately pose their emotional expressions (Hayes-Roth, Ball, Lisetti, Picard, & Stern, 1998). I was therefore interested in whether telling a user that the system might respond to their emotions would lead to a greater degree of emotional expression at the interface. I was also interested in whether users might naturally adapt to systems that respond to emotion by displaying more expression, even when not specifically aware that the system might adapt.

It certainly appeared from the subjective data that participants believed that they were showing more emotion in the conditions where they had been told the system would respond to this. The highest self-ratings of emotional expression were from those participants who were expecting an affective response, but where the system did not adapt. These results suggest that users of an affective system will exaggerate their emotional expression when they expect an affective response, and that this effect will be particularly pronounced where the affect recognition does not appear to be working. This finding duplicates the hyper-articulation effect observed in error correction with speech recognition systems (e.g. Oviatt, MacEachern & Levow, 1998). Interestingly the self ratings for use of emotional expressions (see appendix E for questions and section 5.3.6 for presentation of results) were as high for those who did not expect an emotional response as those who did, in the cases where the system did actually respond to emotion and in fact they exhibited more behaviour episodes as time went on. This suggests that participants may adapt to affective systems by showing more emotional expression, even when not specifically expecting an affective response from the system.

Collectively the results from this subjective data are encouraging for the success of emotional recognition in the short term. All humans show emotions and the findings from this study suggest that users will adopt behaviours that have the potential to aid the recognition process. Despite the encouraging subjective results, I have found limited evidence so far for tangible differences in observable behaviour as a result of the experimental manipulations. It is clear, from the video recording data, that participants displayed a large number of affective signals at the interface. However, this appears to be true for all conditions, not just those in which the participants were expecting, or got, an affective response. A main issue emerging from this large body of data is the complexity of analysis possible and the time it takes to carry out.

Of the observational data that has been coded and analysed only one significant effect was observed, with blink rate found to vary according to whether participants expect the system to be able to respond to expressed emotion or not. Blink rate was significantly higher in conditions where participants were expecting an affective response. As blink rates have been linked to arousal and stress, these results suggest that the expectation of using an affective system may cause participants to feel more vulnerable or may add emotional load (Givens, 2009).

I collected a large amount of data that has only been sampled in this experiment. In-depth coding of emotional expression is ongoing and highly time consuming. It may be that more experimental effects will be found on expression, valence and intensity if more data is coded at a micro-level. The lack of many clear influences of experimental condition on macro-level expressive behaviour, despite the significant subjective results, suggests that participants may to some extent misjudge the degree to which they display emotions when interacting with computer systems. There is some support for this contention from a minority of participants who stated categorically that, once told the system would be ‘looking at them’, they had decided to show no emotion, although, in fact, these participants showed many recognisable affective behaviours during interaction.

A possible limitation of this work lies in the use of the ‘Wizard of Oz’ technique in order to simulate affective responses. The experimental manipulation relies upon the experimenter making appropriate subjective judgments about when the system should make its affective interventions. The Wizard’s responses to participants followed natural ‘gut’ feelings about when participants’ behaviour indicated that they might want a clue. This involved reacting to a very wide range of signals, e.g. providing a clue when participants appeared to display frustration or anger, by tapping their fingers and making ‘tutting’ noises or when participants appeared to display boredom by sitting back and increasingly looking away from the screen. Post hoc validation of the intervention protocols would strengthen the findings and so this was planned as a follow-on study.

Qualitative data added some rich insights into views of participants. The remarks of the participants showed more negative than positive valence, for example describing difficulties with the system. The wording in the initial consent sheet and explanatory notes may have influenced this. The briefings had to be carefully worded, so as not to lie to the participants, but also so as not to reveal the Wizard of Oz aspects of the experiment. They included a phrase saying the purpose of the experiment was to "explore how usable people find computer systems". Overall two thirds of participants were in favour of development of affective systems, despite some caveats, and with none strongly against it.

However, the subjective data from experimental participants (see Appendix G.) described in section 5.3.6, provides encouraging preliminary support for the validity of the experimental manipulation, with participants' post trial belief in the affective nature of the system varying by condition in the ways that one would predict.

5.5. Summary

The user centred approach, empirical methods, and iterative experimental and factorial design, in which I varied the participants appraisal of whether the system was affective and varied system responses to emotional displays, provided a richness of doing and resulted in an interesting data set. The design of the simulated affective system that appeared to detect emotional expressions and react accordingly, when compared to the standard system yielded rich data from participants. Detailed analysis of the observational data has the potential to further inform future design work, e.g. future systems using natural language processing (NLP) techniques for automatic detection of emotion will need to learn from observational studies.

I proved my hypotheses:

- If the system detects emotion, there will be improved performance.

- If the system detects emotion there will be improved satisfaction.
- If the system is affective, then expressions of emotion will increase in valence.
- If the system is affective, then expressions of emotion will increase in intensity.
- If the system is affective, then emotion related behaviours will change.

This work has demonstrated empirically that usability and user experience benefits in both performance and satisfaction can arise from recognising and responding to user affect at the interface. In this experiment, where the response of the system was to offer clues more readily, that response was likely to improve the performance of participants. In learning settings this might or might not be desirable. It is important that clues should be provided appropriately, so as not to remove the challenge of the game, but to create a balance between challenging participants and helping them. As, in this case, I, as the Wizard, had designed the game and had expert knowledge of the challenges it presented, I was confident that I was able to supply clues appropriately for the learning setting. My experimental work has shown that using an affective system or believing a system to respond to affect, causes effects on task performance, subjective satisfaction, ratings of valence and arousal and behaviours exhibited. Following my systematic coding and analysis of data, I found significant gains in performance and satisfaction and identified some interesting issues for further study. My results have potential relevance to the designers of future affective systems. Further research is needed to establish if the benefits found here extend to other applications, platforms, and types of adaptation, and if so, to specify which ones.

The two groups of participants who were told the system might respond to their feelings and so expected the system to adapt to their affective states were rated by the researcher as using emotional expressions with a significantly higher arousal level. These participants also had a significantly higher blink rate. When asked if they thought they had displayed their emotions, these two groups both rated themselves at a significantly high level. Out of these two groups, the group who were told the system would react to affect, but where it did not in fact do so, reported themselves as displaying emotions, highest of all.

Participants' self-ratings as to if they had displayed their emotions was highest in the condition where told the system would react to them. It was highest of all for the group who expected affective reactions but did not get them.

Participants in the condition where the system did appear to adapt to their emotional expressions, performed better and rated themselves as feeling more positive. The researcher perceived them as using more positively rated emotional expressions. Out of the two groups included in this condition, participants in the group told to expect a responsive system and where they did use a responsive system used the most positively rated emotional expressions.

A wide range of affective related behaviours were exhibited. All could be rated for arousal and valence and could be classified as having varied and identifiable modes of communication within them. In the management of the analysis I tried to be as objective as possible, and to validate coding, but coding of the observational data provided challenges and better tools are needed to ensure that coding is systematic and reliable across coders.

The results presented in this chapter have some limitations and lead to a requirement for some post hoc work to validate them. There is a need to explore limitations of the 'Wizard of Oz' techniques and protocols, where manipulation relied upon the experimenter making appropriate subjective judgments about when the system should make its affective interventions. A further experiment could validate the decisions made in my experimental work. In Chapter six I go on to describe the post hoc work arising from this study, and in Chapters seven and eight I discuss the overall outcomes and issues and contributions from the work.

6. Follow-on Experiment

6.1. Introduction

In my main experimental work, as described in Chapters Four and Five, I established benefits of affective systems and that users naturally use a range of non-verbal and emotional signals during human-computer interaction and I raised a number of further questions. This chapter describes post hoc experimental work using a revised response protocol for the Wizard's responses (section 6.2). This follow-on experiment attempted to address some of the issues arising from the experiment described in the previous chapters (chapters four and five). I then go on to discuss results (in 6.3) and discuss the lessons learnt (in 6.4). Section 6.5 concludes by summarising the follow-on experiment.

6.2. Experiment Rationale and Design

The post hoc experimental work built on lessons learnt from the first experiment. I identified a requirement for post hoc work to explore a limitation of the 'Wizard of Oz' technique. In the earlier experiment, the WOZ manipulation relied upon the experimenter making appropriate subjective judgments about when the system should make its affective interventions. This relied on a researcher observing the participants in real time and making the system appear to respond to them. The researcher responsible for manipulating the system did not have a specific protocol to follow, other than to respond to participants' affectively related behaviours (affectemes) and emotional expressions, in real-time, in a 'normal' human way, so that it appeared the system was acting just like a real, empathic person might act. I wanted to validate the decisions made by the researcher manipulating the Wizard of Oz scenario (the Wizard), in responding to participants, by trying to identify the specific cues from the participants, that the Wizard had responded to. I wanted to use those cues as the protocol in my follow on experiment, so as to validate the decisions made in the previous experiment.

This follow-on experiment allowed me to give further consideration to the extent to which users are willing to adapt their use of communication strategies and emotional expressions, when using systems that respond to their affective displays. For this post hoc experiment, I recruited two further groups of participants. I used the same Wizard of Oz methodology and experimental design methods as in the previous experiment, but in this experiment a different protocol was used for responding to user behaviours: I manipulated the system so it appeared to respond to only particular affective related behaviours and emotional expression (affectemes identified as frequent cues in my previous experiment). The independent variable was that the users were, or were not, trained to communicate using the chosen affectemes.

I explain in more detail the research questions and issues that defined this post hoc work and the why and how as to the conception of the further conditions for experimentation. I describe the preparatory work I undertook for this post hoc study, and then the study design, experimental conditions and procedure. This work did not result in any clear contribution in findings but did raise many more questions for research. I discuss the results and issues arising from the study, including the ability and willingness of participants to use specific behaviours.

6.2.1. Research Questions

Shared Codes. In the previous experiment, some users were ambivalent or antagonistic in their attitude towards systems that might recognise all their emotion related behaviours. In terms of user needs, it may be unnecessary to build costly complex systems with comprehensive abilities. Time and resources could be saved if simpler systems would suffice. Instead it may be more agreeable to users and more cost-effective for developers to give the user more control and opportunity to decide what emotional behaviours are monitored by the system, by using agreed and shared codes of behaviour and to concentrate resources on recognising instances of these.

Adaptive Behaviour of Participants Users have quickly embraced wave after wave of digital technologies in the past decade and novel technologies often require adoption of new means of communication or adaptations of old ones. There is evidence that users will adapt their behaviours to new media such as use of emoticons, email styles, and input devices such as Sentoy (Paiva, Andersson et al, 2002). In my main experiment I found that users adapted their behaviour when interacting with an affective system, depending on their appraisal of the system. I wondered if users would be willing to adapt their behaviours in a way prescribed by me for interaction with an affective system. In my main experiment it was apparent that participants were often not very aware of their own use of emotion related and non-verbal behaviours. Those who believed they had shown most emotions did not show use of a significantly higher number of observed affectemes.

6.2.2. Preparatory work

In the experiment described in chapters four and five, the Wizard responded to displays of emotions by sending clues whenever the Wizard felt that there was an appropriate emotional ‘prompt’ or display from a participant. But the Wizard did not follow a particular protocol, as the focus was more concerned with exploring the kinds of displays participants naturally make. On reviewing the experiment I realised that it was also important to analyse the responses of the Wizard and that a further experiment could usefully be carried out where the responses of the Wizard were controlled in some way or where the behaviour of the participants was controlled or refined to some extent.

In order to analyse the behaviour of the Wizard, and to understand what behavioural cues the Wizard responded to, I reviewed the sample video data from the first affective experiment. I examined the behaviours of participants in the 15-second window immediately preceding clues given by the Wizard. Behaviours in these clips was coded using the affecteme coding scheme as described in chapter four and appendix D. I found a whole range of behaviours had prompted clues to be sent. A strength of my coding scheme was that it allowed systematic coding of behaviours, despite some variance in those behaviours (for example the variations on a ‘chin dump’ illustrated in chapter 5.2.) However this method relied on giving precedence during coding to the most significant observed behaviour, while other behaviours could co-occur – for example smiling or frowning during a chin dump. Other variations could occur such as duration, level of arousal or onset / peak / offset time of the behaviour. But these variations make it difficult to establish exactly what the

Wizard was responding to. Sometimes it seemed that an absence of behaviours, for example when a participant was very still for a long period, prompted a clue to be sent. The commonest behaviour identified was complex: a ‘double-take’ consisting of a postural and head shift and a puzzled expression or one of these elements alone, i.e. a postural twirl or nod or shake of head or puzzled expression.

6.2.3. Pilot Behaviour Training

In a small pilot study of three users I experimented with training them to perform these three complex behaviours intentionally as cues for the three conditions. I found this very difficult to achieve, despite experimenting with the use of prompt sheets with textual instructions and still images, spoken instructions, use of video snips and acted behaviours and different levels of training.

Non-verbal behaviours are often unconscious and intentional use of three complex behaviours for three different purposes was just too challenging for training in a short time. Eventually the training was reduced to training for just one complex behaviour. This was the use of a ‘double-take’. Double-takes were realised in a variety of ways, and typically consisted of a postural shift of the body and / or head and a slightly negative, puzzled facial expression. I do not claim to map this movement with any certainty to an underlying emotion.

Double-takes were selected because the primary mode was clearly identified as a postural shift and all my users had shown some spontaneous use of this technique during their overall interaction. All 60 participants used postural shifts at some point in their interaction (duration range 1 – 41 seconds, mean 8.62). This behaviour also seemed to be the most intuitive for participants to learn in my pilot training sessions. Responding to double-takes gave the Wizard a slightly wider remit than a postural twirl as double-takes could be realised by facial expressions, as well as head and body movements.

In pilot sessions, training to use a ‘double-take’ was attempted. This proved more difficult than expected. Using static photos of the desired emotional expression was unsatisfactory. Next I tried annotating the static photos with arrows to indicate movement. I hoped that a comic strip or short video of the desired behaviour approach might prove a slightly easier tool to communicate with, but with more detail or moving images, participants did not seem able to pick up all the three elements of the double-take – the postural twirl, head shift and puzzled expression. Eventually the best method proved to be a ‘live’ demonstration by the Wizard, with three repeated efforts to stress each of the three aspects.

A training protocol was arrived at. The most successful training method was using a set script to describe a double-take accompanied by an acted demonstration of a double-take by the trainer. The participant was then required to demonstrate a double-take. In case of inability to perform a double-take, the training was repeated for up to five minutes (see discussion in 6.4).

6.2.4. Study design and procedure

The study used in the main experimental work was replicated with two groups of participants with two alterations to my study design.

Firstly I selected one independent variable. The individual participants from one of my groups received five minutes training in the use of a ‘double-take’, immediately prior to their interaction with the system. They were told that this behaviour would be recognised by the affective system (as they believed it to be) and that the behaviour would prompt the affective system to supply them with an extra clue. In the case of the other group they were simply told that the system might respond to their emotional expressions.

Secondly the Wizard was limited to use a set protocol to decide when to supply extra clues and only to supply an extra clue if a participant executed a ‘double-take’.

Experimental conditions In the main experimental work the researcher attempted to respond to any emotional expressions, as if interacting person to person (see briefing scripts in Appendix E Participant Information). In this follow-on experiment, the researcher followed a strictly limited protocol. The only emotional expressions responded to were variations of the ‘double-take’ affecteme. This could consist of a postural twirl, a head shift, a puzzled expression or any combination of these. The Wizard's intended response to any of these behaviours was to supply a clue in response.

I was interested to see if users would adapt their behaviour to use double-takes, and whether they would intuitively increase their use of this emotional expression, or if their use would be increased for the group given specific training and instruction.

Procedure This experiment used the same controlled laboratory conditions and problem solving game as the main experimental work. Two further experimental groups of 15 participants were recruited, paid and their consent for participation was gained. Again the WOZ approach, Word Ladder game, networked computers and two way mirrors were used to simulate a responsive affective system. Performance and demographic data was collected in the same way.

One group was briefed exactly as the group in the told affective condition in the first experiment. They were told they would be interacting with a system that might recognise some of their emotional expressions. The other group was similarly told they would be interacting with a system that might recognise some of their emotional expressions and in addition told that they would be asked to use some particular emotional expressions (explained and demonstrated by the researcher) during the game (see Appendix E). The plan was that for both groups the Wizard would respond only to double-takes.

Each participant was given 5 minutes 'live' training on production of a double-take and once they were successfully able to produce an example of this behaviour, they were told that the system would recognise and respond to their double-take behaviours.

6.3. Results

The participants' arousal was measured, following the interaction, by rating their affectemes during the one-minute sample video clips, using a scale based on the Self Assessment Manikin, (where 1 is very high arousal and 9 is very low arousal). When compared to the control group from the first experiment, the overall mean rating was 4.57, SD 0.60. For the different conditions the means and standard deviations were as shown in Table 18.

The participants' valence was measured, following the interaction, by the researchers rating the affectemes during the one-minute sample video clips, using a scale based on the Self Assessment Manikin (where 1 is very positive and 9 is very negative). When compared to the control group from the first experiment, the overall mean rating was 5.20, SD 0.70. For the different conditions the means and standard deviations were as shown in Table 18.

Group 1 (from experiment one)	**Group 5**	**Group 6**	**Overall**
Participant not trained	**Participant trained**	**Participant not trained**	
Wizard responds intuitively	**Wizard responds to double-takes**	**Wizard responds to double-takes**	
Observed Arousal Mean 4.54 SD 0.55	Observed Arousal Mean 4.48 SD 0.68 (highest arousal)	Observed Arousal Mean 4.72 SD 0.59 (lowest arousal)	Observed Arousal Mean 4.57 SD 0.60
Observed Valence Mean 4.70 SD 0.43 (most positive)	Observed Valence Mean 5.54 SD 0.47 (most negative)	Observed Valence Mean 5.35 SD 0.70	Observed Valence Mean 5.20 SD 0.70
Affectemes Mean 7.40 SD 3.1 (most affectemes)	Affectemes Mean 6.40 SD1.54	Affectemes Mean 6.33 SD 2.22 (least affectemes)	Affectemes Mean 6.71 SD 2.38

Table 18. Results summary

6.4. Discussion

There were all sorts of difficulties during this experiment at every stage and altering the response protocol gave me no valid statistical findings. However it did raise a number of interesting issues.

In the main experiment, my participants significantly and intuitively adapted their behaviour in the various conditions, but when, in my revised response protocol experimental work, it came to intentional, trained adoption of behaviour, there was a mismatch and intentionally altering their non-verbal behaviour proved very difficult for users to manage.

Firstly my participants found learning a complex non-verbal behaviour very challenging. In some cases I did not managed to elicit a double-take within the five minute training. I learnt that there are significant challenges in training non verbal behaviours, and that I need further experimentation with training methods, a longer time period for training and a greater number of practice sessions to properly train participants in a complex non-verbal behaviour task, involving emotion related behaviours such as this.

There were individual variations in non-verbal skills, with two of the fifteen very quickly understanding what was required and making recognisable attempts to comply both in training and during the game play. Five appeared to be absolutely lost during the training and made no effort to use double-takes in game play. The remaining eight seemed to have some grasp of the requirements and successfully executed the double-take in training but seemed to forget about it during game play.

Once immersed in the task, trained participants in any case often seemed to forget about using double-takes and were no more likely to use them than the control group. There seemed to be several factors at play here. Firstly non-verbal communication is often not consciously performed and requires a rather unnatural cognitive effort. Secondly the task itself provided a significant cognitive load and this may have interfered with the intentional execution of double-takes. Thirdly I was dealing with personal styles and preferences. Double-takes are a complex behaviour with several elements to it and the demonstrations I used might not have matched to the normal range of behaviours used by that individual. For one person a double-take might mean more of a postural twirl and less of a facial expression. For another person more of a head twirl and so on. The trainer was not sure what to stress and participants were not sure what to perform. This uncertainty may have been compounded if the demonstrated behaviour did not match to their personal style, or if they were not very skilled in or aware of non-verbal behaviours.

On the other side of the WOZ scenario I also encountered difficulties in noting and responding to a set behaviour. The Wizard had a considerable workload when using the original protocol. This post hoc work added two tasks: firstly training sessions and a secondly a requirement to notice and respond only to double-takes. In planning these sessions, this sounded like a simplification of the management required for the first experimental tasks. In this iteration, half the subjects might be expected to behave as they had been trained to, in a more predictable way. With both groups of participants, the Wizard would respond to just one, specified behaviour. In addition, at the outset of this follow on experiment, the Wizard already had the benefit of considerable experience from running nearly 70 trials (for the previous 4 groups of 15 subjects, plus pilots and rejects).

The first task for the Wizard was to introduce the participants to the scenario and train them. As the experiment began the Wizard had to set off various recording devices and quickly move to the control room. Throughout the task, the Wizard had to monitor two screens, one showing the user behaviour and one the interface for the control of clues that could be sent to the participant. The Wizard also had to monitor the behaviour of the participant via the two way mirror in real time, ignoring some non verbal messages that they might naturally have responded to, and sending clues only in response to double-takes.

It proved very challenging and the results of the experiment were confounded by the problems. When I reviewed the data I found that participants had sometimes performed something that might be construed as a double-take that the I, as the

Wizard had overlooked or ignored, while in other cases a partial instance of one element of a double-take provoked a response from the Wizard and a clue was sent.

In the first experiment, the WOZ manipulation task was already complex. The researcher had become practiced in the ways the first experiment required manipulation to facilitate the scenario. Changes in the protocol required unlearning previous behaviour and that proved difficult. The newly revised task was just too complex and with hindsight I should have involved an extra person to facilitate it. The stress of the Wizard may have been contagious and contaminated the participants' emotional states.

There was also a very fundamental error in my overall design that I did not identify until after the event. This was the 'help' button. The game interface I designed had an on-screen 'help' button (as games typically do). Participants could hit this button at any time to get a clue, which would appear in a new window. How much they used this depended on personal style and the game play strategy they adopted. For example some players preferred to 'beat the system' by dint of their own cognitive efforts and hated to use it, struggling on for a long while and only hitting 'help' as a last resort. For others it was obviously more important to beat the clock by any means at their disposal and they would hit 'help' repeatedly and quickly open as many clues as possible, minimising the windows so they could view them all at once and so stood a better chance of synthesising information from the various clues, in order to increase likelihood of reaching the correct answer, more quickly. Computer literacy also played a part as although all my participants were deemed to be computer literate, some found it difficult to manage more than one window and had obvious difficulty navigating between windows once a clue window appeared. There were also clear differences between individuals as to preferences for mouse and keyboard use.

In the previous main experimental study, the help button did not seem especially relevant. Participants were told that the system might respond to some of their emotional expressions, but they were given no expectations that their use of emotional expressions would prompt clues or help them in any way.

My very significant issue in this post hoc work was having the help button in place in this iteration of the experiment. One group of participants were told that a particular non-verbal behaviour would prompt a clue. They could also prompt a clue by hitting the help button. With hindsight, it is not surprising that faced with the choice between the familiar task of simply hitting a help button, or of performing a recently learnt and very complex non-verbal behaviour – most users will hit the button. Alternatively a more easily usable system could equally well be designed with single word recognition so participants could simply ask for help when they wanted it by saying ‘help’ or ‘clue’.

I did not feel this work to explore using a response protocol for a particular non verbal expression was a success, but it did illustrate some very important issues for design and evaluation of affective computing systems and help to identify a number of areas for follow on work:

1. Affective systems do have potential to improve both the performance and the satisfaction of users. Affective behaviour is often unconscious and difficult for users to communicate about verbally. An attempt to bring affective behaviour into the conscious realm, by requesting intentional behaviours, interferes with the affective process.
2. It might be expected that recognition of one non-verbal cue might be equivalent of single word recognition and so quite easy to perform. However the non-verbal behaviour selected was quite complex with several elements and it was often quite difficult to decide when in rapid sequences of behaviours, often occurring simultaneously and expressed in different modes, the behaviour could properly be ‘counted’ as a double-take. More training of the Wizard, or a Wizard with no other tasks, or selection of a less complex non-verbal behaviour could be tried.
3. It is never wise to over complicate system design. At worst, it may be completely outside the normal range of conscious control of emotional expression for participants to alter their emotional expressions, or at best participants’ skills in emotional management are likely to vary widely. If hitting a ‘help’ button is a well tried, tested, usable, liked feature, then only something even more usable and liked will motivate users to change their

habit. In the previous experiment I showed that a system with affective responses can help users, but it must be intuitive so as not to add to their workload. It might be more useful to explore direct intentional intuitive ways for people to communicate how they feel, rather than trying to make them adapt their behaviour via training. Non-affective means such as responding to the word 'help' should be considered.

4. Affective systems need to build on the intuitive skills of participants, and not to confound them. It might be better, using prototypes during iterative phases of participative design, to observe any intuitive adaptations and capitalise on them, rather than trying to specify necessary behaviours.
5. There are issues about training participants for affective behaviours. My training was not always very successful and the taught emotional expression skills were not generally sustained during game play. Individual differences of participants such as personality, mood or culture might all impact on this (as described in previous chapter 2.3.2).
6. Following training sessions, particularly those that went less well, the researcher reported feeling very flustered and negative which may have influenced performance on identifying and responding to double-takes. Undoubtedly the same thing may have occurred for participants, although I did not collect direct reports of this. There is more work to be done on understanding the carry-over and contagion of longer-term feelings to current behaviours and their effect on performance and satisfaction in such experimental settings.
7. There are issues around personalisation of non-verbal behaviour. The post structural approach – that involved spending considerable time looking for patterns in my data and reflecting on behaviours observed – led to reflections on how it is the differences in behaviour that count and awareness of the different behaviour patterns of different individuals that prompted changes in the behaviour of the Wizard. It seemed that what prompted the Wizard to send a clue, was often not related to the *type* of behaviour (for example whether a double-take or a laugh), but more whether a participant *altered* their pattern of behaviour (for example suddenly moving around or suddenly keeping still). There is scope for much more sophisticated sequential analysis of my data and

data mining to explore sequences and temporal patterns of the various behavioural events in different modes, their meanings in different contexts and best means to respond to them.

8. There are issues around personal non-verbal styles, highlighted by the failure of this post hoc experiment. It is recognised that people have different preferred verbal communication styles e.g. assertive, passive etc. Most people use a range of styles depending on context, for example depending on the locus of power or rules of politeness (Brown & Levinson, 1987). For example, it might be appropriate to be more assertive as a teacher and more passive as a student. The appropriateness of the degree of assertiveness or passivity might vary with context, for example in some Eastern cultures it is considered appropriate for students to show extreme passivity such as not making eye contact with professors, or not smiling at them. It seemed my participants used a range of non-verbal styles that they could quite intuitively switch between.
9. There are issues about structure of responses to non-verbal behaviours. I spent a lot of time building my coding scheme to categorise observed behaviours, but I feel there is follow on work needed on how behaviours are appraised, (as they were in practice by the Wizard), using a structured approach. Valence and arousal ratings within the affecteme coding scheme provided a strong starting point, but there are many more aspects, such as onset, offset and duration of different aspects of affectemes, or dominance of modes to be explored.
10. There is a real need for tools that speed up or add to the ease of manual coding, as this is still frequently required and is a very time and labour intensive process. It is likely that it will be necessary to manually code behaviours for some time, as it will be necessary to manually code in order to verify the performance of early systems and to do this until systems are proved reliable. The issues around reliability, variability and training of coders requires further and deeper consideration.

The original intention was to compare key measures from these two groups with a control group, which has no intervention (using the group from the first experiment, who were not told the system was affective and for whom the system did not act affectively). I intended to report post hoc comparisons. However it was felt that the behaviours exhibited by the participants and the subsequent responses of the researcher were too compromised to make statistical inference meaningful.

6.5. Summary

Based on issues arising in my main experimental work, I carried out a post hoc experiment with a revised protocol for responding to the behaviour of participants. In this iteration of my controlled experiment, the independent variable was whether the participants were trained in using a non-verbal signal (a double-take) and told that using it would prompt a new clue to help them in the game.

I had difficulties with many aspects of this study. Both the participants and myself (as the WOZ facilitator) were extremely challenged by the task of consciously using non-verbal signals. Although I did not get any statistically significant results, I raised a lot of issues for further consideration. These include implications for development and design of affective systems and issues where further exploration of affect is needed: the need for simple intuitive systems; issues around training non verbal behaviour; carry-over of emotional feelings; individual differences and non-verbal communication styles; a need for a structured approach for understanding appraisal of behaviours and a need for more tools to support coding and other processes in affective system design and evaluation.

In Chapter Seven I will discuss issues arising from my work and in Chapter Eight, I draw some conclusions arising from the work, describe the contributions it makes and consider opportunities for follow on work.

7. Discussion

7.1. Introduction

In my main experimental work (as described in Chapters Four and Five) I established some benefits of affective systems and found that users naturally use a range of non verbal and emotional signals during human-computer interaction. The initial experimental work, along with the follow-on experiment (described in Chapter 6) illustrate some lessons learnt and raised a number of issues for discussion that require consideration and debate. In the following sections I discuss the complexity of emotion related behaviours (7.2), the difficulty in distinguishing conversational and emotional behaviours (7.3), issues around training non-verbal behaviours (7.4) and emotional contagion (7.5). I comment on study design (7.6) and coding issues (7.7). Finally in Section 7.8 I discuss implications for the design of future affective systems. The discussion is summarised in 7.9.

7.2. Complexity of Emotion Related Behaviours

From closely reviewing my data and observing the behaviour and responses of the researcher acting as the WOZ controller, I can state that the basis of responses is very complex. Using technology is an interactive process and emotions change dynamically during interaction according to variables, such as success in using an application. Secondary or more subtle shades of complex emotions may be identifiable or confounding. Issues so far identified for further exploration include:

1. Idiosyncracies of participants. The unpredictability of humans and the multimodal nature of communication creates difficulties for any automatic detection system that sets out to interpret emotional expressions without situational awareness. For example a number of my participants obscured some of their facial actions with their hairstyles or by wearing hats or glasses, chewing gum or peering partly off camera at times. This is an on-going, real world issue for future systems.
2. Angle of view. When the same expression on a face is seen from different angles it can appear to be more smiling or unhappy to a greater or lesser

degree, depending on the degree of angle (seen from above, faces appear to smile and vice verca). Positioning of cameras is an important consideration.

3. Mixed messages. These occur between modes and sometimes are used as a communication strategy, for example irony is often conveyed by a mismatch between tone of voice and choice of words. Some facial actions can be used for different purposes. For example there is a typical expression of grief – known as ‘Darwin’s grief muscle’ as a result of Darwin’s book (Darwin 1872) on the expression of emotion. Bakeman and Gottman (1997) noted that this is the exact expression typically used by the comic actor Woody Allan to signal delivery of a joke. It is not possible to accurately assess the emotional expression unless you make an appraisal of the intent behind the emotional expression. Interaction of all spoken and non-verbal modes need consideration.
4. Duplication of meaning / redundancy. Duplication of meaning in multiple modes occurs and is know as ‘redundancy’. It is a useful strategy in situations where some modes may be lost or poorly perceived, such as using gestures to assist understanding in a noisy environment, or tone of voice to understand valence when face is not visible. Modes are not only used to communicate affect, but can add emphasis or extra information, for example Roberts and Bavelas (1996) describes a participant’s use of gesture in a narrative about a dynamic event (a ‘near-miss’ car crash), where spoken information is enhanced by gestures to symbolize movements and add information about direction, position and force and the consequent intensity of the physical and emotional experience.
5. Communication of stance. Emotional expressions may also be used to convey an attitude or judgement rather than an underlying feeling, for example a frown to show disapproval rather than real anger. Emotional expressions can be at odds with underlying emotional experience, for example a smile faked for personal or social reasons e.g. to greet someone. Inhibition can speak, and in certain contexts a blank face can give a powerful message (e.g. where animation is expected). Emotional expressions may be under conscious or unconscious control and may be ‘leaked’ unwittingly. Some researchers use theories that link emotion-antecedent cognitive appraisal and associated action tendencies, the expression of which we may inhibit or regulate, for example

Kaiser, Wherle and Schmidt (1998). Roberts and Bavelas (1996) suggest that many widely reported beliefs about the meanings of non-verbal communication are based on very flimsy or non-existent evidence. A good example of this is an oft reported adage used by many commercial training companies, that communication is 93% non-verbal. This is attributed to Albert Mehrabian, a professor of social psychology, who conducted a study in the 1960s that found in a particular context and for a particular group of participants, carrying out a particular task, 55% of the communication was non verbal and 38% tone of voice and only 7% verbal. He never intended this to be generalised to other situations and now describes how he cringes when he hears this information offered as a general fact and misattributed to his findings (Mehrabian & Wiener, 1967; Mehrabian & Ferris, 1967).

6. Relationship of meaning to mode. It may be that certain single modes, such as facial expression or gesture, are highly significant indicators for certain emotions, but if so we have not yet reliably mapped them. In the absence of sure indicators, we must consider information from all modes (both spoken and non-verbal). In face-to-face situations, we constantly make judgements about people's affect by using all of our senses, and our situational knowledge of place, time and shared experiences.
7. Episodes of behaviours relating to affect (affectemes). I developed this system to enable coding of affective behaviours at different levels of detail. It proved useful in this experiment, but requires further use and exploration by a range of researchers and with a range of issues, in order to validate the approach.
8. Onset and offsets and peaks of episodes. It is likely that just as variation in the onset and offset and peaks of sounds in words alters meaning, so the onset, offset and peaks of aspects of emotional expressions will alter meaning. This might apply across the episodes that coders identify as discrete behaviours. For example in the 'double take' identified as a trigger for the Wizard to send a clue – it would be interesting to see if it is possible to identify exactly when the 'double take initiated, when it was perceived to have ended and when it was perceived to peak and at what point in that cycle the Wizard responded to it.
9. Multiple modes (body movements, speech, facial expressions etc.) used within episodes. The affectemes coding system did consider multiple modes of

communication of emotional expressions. However the coding system relied on coders noting and segmenting the most prominent behaviours as discrete episodes, and then looked for other components that made up that episode. It would be interesting to look in more detail at patterns within episodes and to carry out data mining of coded segments to look for patterns of behaviours and in particular sequential patterns.

10. Onset, offsets and peaks of modal behaviours. The complex episodes of emotional behaviours were made up of any number of modal behaviours, For example the ‘double take’ included elements such as a postural twirl, a shift of the head, a frown, but it could also include an affect burst or a smile or any number of other elements. The exact onset and offset of each element could be timed and analysed further. This might be helpful to developers of embodied conversation agents (ECAs) who strive to recreate more realistic characters. It is important to understand the timing of elements as well as the interplay of different elements. For example, a slow smile versus a fast one or one that peaks in conjunction with placing of stress of a syllable of a certain word. My analysis did not time these aspects, but it would certainly be of interest to do so.

11. Dominance and subordination of modes used within an episode. It would be interesting to explore in more detail how and why coders made their decisions about discrete episodes of behaviour. The agreement on some episode boundaries was quite striking, but some coders perceived more subordinate episodes than others. It might be that some coders were paying more attention to some modes of behaviour than to others.

12. Identification of baseline behaviour for a subject (i.e. what was normal or typical behaviour for the person and how long was needed to establish that). The design of the experiment allowed the Wizard to spend some time (usually about half an hour) interacting with participants and noting their communication styles. We automatically tune in to our conversation partners and start adapting to them within moments of the start of any interaction. We are soon aware of how expressive they are and of any idiosyncrasies in their behaviour patterns, including the way they use emotional expressions. It would be interesting to explore how coding might vary, depending on how

much exposure a coder has to any one individual. It might be that coding varies over time, or that initial episodes of coding need to be discarded, until enough time has elapsed for a coder to have 'tuned in' to an individual under observation. The length of time required for tuning in might vary between coders. This all requires further exploration.

13. Deviations from baseline behaviour norms. Once a coder has tuned in to the behaviour of an individual under observation, then any deviations from the norm are likely to be significant. So it might be that a very small smile is rated as having high arousal for an individual who rarely smiles, while for another individual who smiles frequently, a small smile might only be rated as having low arousal. Understanding this kind of variation is intuitive for people, but needs exploration in the domain of affective computing. Recognition systems might need to establish baselines before starting to analyse behaviours in any detail.
14. Variations in skills in using emotional expressions. Individuals vary in their skill for unconscious or conscious displays of emotional expressions to communicate how they feel, and they vary in emotional management or emotional IQ skills that might include suppressing displays of emotion. We need to understand this variation and what it might mean for emotion recognition in computing.
15. Variations in skills in detecting and interpreting emotional expressions. In the same way as individuals vary in display skills, coders are likely to vary in their skill in both recognising and interpreting emotional displays. We need a means of screening coders for their sensitivity to and understanding of emotional displays, so that we can be more confident in coding procedures.
16. Variations in skill in responding to emotional expressions. Systems that make or simulate affective responses to emotional displays will need to use knowledge of the social rules for use of these as appropriate for the given interaction. The knowledge of social rules for engagement and skills in responding appropriately varies between individuals. This is another area it would be interesting to explore, for example by repeating the experiment using different Wizards to explore how their responses vary.

17. Assessing the extent to which behaviours were intended to convey emotion or to supplement communication. This experiment relied on the Wizard and on coders deciding if behaviours were intended to convey or to communicate emotion. A small subset of participants took part in retrospective walkthroughs to explore their own interpretation of their own behaviours. It would be interesting to extend this to all participants, or to use probes during the experiment or to devise other methods to try and understand more about when participants wanted their emotional behaviours to be recognised, understood and responded to.
18. Assessing the extent to which behaviours showed emotional messages that were unintended, underlying, or of which the participants were unaware (akin to Ekman's 'leakage' (1982). There were notable instances within episodes of fleeting expressions that were reminiscent of the leakage of emotions described by Ekman, but rejected by Bavelas, Black et al (1990). It would be interesting to explore these further and to explore with participants whether they were a part of their intentional behaviour or not.
19. Assessing if behaviours related to a short-term emotional event (such as to the interaction in hand) or if they related to longer term emotional constructs, (such as mood or personality traits). Personality of participants in the experiment was assessed, along with affective state following the experiment, but affective state prior to the experiment was not assessed and longer-term mood was not assessed. It would be interesting to repeat the experiment taking more measures into account and to consider the interplay of these variables.
20. The extent to which the presence or mood of the manipulator proved contagious to the subject. No measures were taken of variables such as the mood, affective state or personality traits of the Wizard or of coders. The state of the Wizard might have varied on different occasions and so have affected the participants, and the mood or psychological state of coders might also be significant to their coding skills. It would be interesting to develop tools and measures for these variables. I noticed differences in behaviour early and late in the interactions (when the subject has just interacted or was about to interact with the facilitator) and consequently discarded the first and last minutes of my data.

21. Assessment of coders and training and reliability of coders is an issue that needs further exploration and consideration. Individual differences in communicative awareness and skill may be linked to factors such as gender, mood or personality issues.

Further work and experimentation is needed to explore all these issues that are critical for development of affective systems.

I particularly note that most work in this domain has focused hugely on non verbal behaviour and that integration with spoken and written language is much needed. Words are a strong component of both communication and of emotional interaction. I have already successfully used some methods borrowed from the social sciences and linguistics, such as a coding system based on phonetics and observational analysis based on the methods described by Bakeman and Gottman (1997). Discourse analysis has long been successfully adapted to look at turn taking in human-computer interaction (Norman & Thomas, 1991). Systemic Functional Linguistics has potential to help inform human computer interaction analysis and the Appraisal Analysis Framework in particular may be useful (Martin & White, 2005). I hope to carry out future work using a Systemic Functional Linguistics approach and in particular to apply the Appraisal Analysis Framework and develop it to embrace non verbal as well as verbal aspects when exploring interactions.

7.3. Intentional Affect or 'Plain' Communication?

Among the complexities of affective related behaviours, is the issue of whether it is possible or desirable or necessary to identify when a behaviour is affective, when it is related to an emotional feeling, when it conveys an expression of emotion, or when it is 'simply' part of the communicative content of the interaction. On one hand it could be said that everything relates to affect, but counter arguments (e.g. Roberts & Bavelas, 1996) suggest that conveying emotional states is not always or often the primary function of behaviours. My analysis suggests that many behaviour episodes exhibited by participants are related primarily to management of turn taking, and judgements of the game, rather than to 'pure' visceral affect, either intentional or unintentional. For example in conversation we often nod or smile to encourage the other speaker and affirm what they are saying and shift our posture when we are ready to take a turn in speaking. Participants exhibited these sorts of behaviours during game play, and so it seems likely that these non verbal behaviours might be more to do with social management or turn taking rather than intentional or non-intentional revelation of inner feelings. More work is needed to understand this and I hope to return to this issue in future work.

7.4. Training Emotional Behaviours

I have identified that behaviours relating to affect are complex. My research considered how users' emotional expressions might vary when they use an affective system. It has been suggested that, in the short term at least, emotion recognition technology will need users to exaggerate or deliberately pose their emotional expressions (Hayes-Roth, Ball, Lisetti, Picard & Stern, 1998). I was therefore interested in whether telling a user that the system might respond to their emotions would lead to a greater degree of emotional expression at the interface. I was also interested in whether users might naturally adapt to systems that respond to emotion by displaying more expression, even when not specifically aware that the system might adapt.

I found that users clearly did adapt their behaviour, either when given different information, (if told the system was affective or not) or when given a different interaction experience (if WOZ system responded to affective related behaviours displays or not). In either case users' appraisal of the situation was altered and they adapted their behaviour accordingly. Interestingly they appeared to be trying to repair the interaction by increasing their use of emotional expressions when they expected the system to respond to them, but it failed to do so. They seemed to have some degree of conscious awareness of this process, as they reported that they felt they had showed more emotion in this case.

However when I set out to train subjects to use particular emotional expressions I encountered difficulties. Some of the issues are centred around choices of training methods. Some issues are concerned with conscious and unconscious and intentional and unintentional use of modes and messages in interaction. The Hayes-Roth, Ball et al (1998) vision of requiring users to pose emotions may not be feasible.

7.5. Feelings Confounding Feelings

The behaviours I sought to study may well have been altered either by my study, the procedure, the interaction, the observation or the attempts at evaluation. Spoken communication is intentional but processing is automatic and we are usually unaware of it. This is arguably even more so for non-verbal communication. Drawing awareness to the process in any way, by training attempts, by discussion, by observation or by evaluation, may disrupt and corrupt the very thing we are trying to study. Contagion of emotional feelings can occur between people, as can mirroring of emotional expressions (e.g. yawning that is even 'caught' by non humans such as dogs). There may be knock on effects from interactions and we do not know how long these last. I noticed differences in behaviour early and late in the interactions (when the subject has just interacted or was about to interact with the facilitator) and consequently discarded the first and last minutes of my data, as described in 4.2.9). We do not have firm evidence of how long these primacy and recency effects last or their variation. Studies need to consider this issue and design of studies needs to be very rigorous so as not to be confounded.

I found a correlation between being part of the group who were told the system was affective and reported liking of the system. This might indicate that a change in appraisal led to an altered degree of experience.

7.6. Study Design Issues

7.6.1. Preparatory work

Piloting. Studies need considerable piloting if they are to succeed, something I failed to do this sufficiently in my experiment using a revised protocol. I fallaciously assumed that an iteration of the previous experiment with a revised protocol would be easier than the original experiment. I ploughed on despite warning signs and it is important to be prepared to abandon, revise and reiterate study designs as required in order to make significant findings.

Resources and design. There are also issues about manpower and man hour resources required for studies that are intensive, both during procedures and analysis. I set out to observe a large number of variables via use of scales and observations, and this made the procedure rather lengthy. I found it quite hard to balance the desire to observe variables and the need to focus on the most critical variables. This is partly due to the complexity of the domain. Careful design of the study is needed to manage this. Clear decisions are needed about detail of coding required to answer the research question, as this method of ever more detailed coding is an open ended task.

Experimental design. The 2x2 design rules out concerns about the influence of the observer on the participants. A Hawthorne Effect would have showed up for all four conditions (but it didn't), and the Told-Standard condition rules out the possibility that it was the clues alone that were responsible, since the Told-Affective condition fared better.

7.6.2. Subjective reports

Showing emotion. It was notable that the vast majority of participants believed that they had shown emotions in some way and they described a number of ways that they had shown emotion, showing considerable insight into their own behaviours. Those in the 'not told affective' condition had least to say about their emotional behaviours they had shown and those in the control group (not told affective and standard application) said least of all. A number of participants in the 'told' affective' condition said they tried not to show emotion or in one case, that they gave up trying to show any.

Game adaptation. Asked if the game had adapted to them, in any way, around one third of the participants agreed, with the comments they made split between those who thought adaptation was made by the game and those who thought that they themselves had adapted their behaviour. A number made comments to show they ascribed the simplicity of the game design as affording their adaptation. The group in the control condition, who were not told to expect adaptation had least belief and made least comments on adaptation. The two groups told to expect an affective system made most comments and from those two groups, those for whom it did react with apparent affect made most comments of all.

Despite the lack of empirical findings from my follow on iteration of the experiment, I found that the experiment using a revised response protocol, did give some benefits. I amassed a significant additional amount of data and having to struggle to understand the story held within the data, led to deeper insights and understandings and identified further important issues and I learnt important lessons from it.

I already had a strong focus on user-centred iterative design and this work has reinforced my commitment to this and led me to explore participatory design with users in subsequent studies.

7.6.3. Using a WOZ Scenario

My WOZ scenario was successful in many ways. It did allow me to carry out and control my experimental design. My participants expressed belief in the capabilities of the affective system, which was surprising at times and despite an overt camera and two way mirror that might have been expected to cause them to speculate about the veracity of the affective system. For example, one subject, for whom the system acted affectively, asserted at the end of the experiment, “I know how you did that”. By chance, the case of the PC tower I was using had been repaired at some point and had a small Perspex panel in its fascia. He went on to explain his conception that the PC was monitoring him and responding to him via this Perspex panel and that he believed the Perspex panel to be a futuristic emotion detector.

I did have issues of controlling the scenario when I revised the protocol, which I hope to address in the future. Partly because of these issues, I have so far failed to validate the Wizard’s responses and I need to carry out further work to explore this. This has raised issues around consistency and what constitutes consistency in this kind of interaction. Both in my review of literature and in my own work, I have identified many issues around complexity of interaction. It may be that reducing responses to a set protocol that does not reflect that complexity is to oversimplify and fail. One way forward might be using participatory design to understand what signals users are able and willing to show or what responses they are able and willing to use.

7.6.4. Baseline Behaviours

I note that the WOZ controller responded to changes in the subject's behaviour. My procedure required the controller to spend some time with the participant, before the experiment started (for pre test questions etc.) and it seemed that this, and / or initial observations during the experimental trial, created an understanding of baseline behaviour and expectations of a pattern of behaviour. It was often when the subject departed from these that a response was initiated and a clue given, rather than in response to any particular behaviour on the part of the participant.

Speech recognition systems can be speaker independent or speaker dependent and systems are found to be improved when there is an element of training for the system to learn to recognise the way a user speaks. Hence training for recognition of the user's voice is often an option on speaker recognition systems (e.g. voice controlled in-car satellite navigation systems). It may be that some similar type of system training for recognising certain aspects of affective behaviours could be useful in affective computing.

I speculate that the introductory session may have facilitated appropriate responses, or may have been a confounding variable. I would like to carry out more trials to explore this and to explore issues such as how long the Wizard requires exposure to an individual to understand their baseline.

7.6.5. Normality of participants

I found it difficult to recruit the relatively large numbers of subjects needed for sufficient trials in my experimental design. As with many studies, I resorted to using members of the student body as participants. The eminent psychologist George Miller and his colleagues as long ago as the 1960s estimated that most experiments have used college students (who are described as accessible and 'cheap' and also better educated than the general population) and that there is a real danger that results from such experiments cannot be generalised to other domains, or applied to the general public (Dyer, 2006). I collected some demographic data on my subjects that showed they were not representative of the general population (for example a high representation of middle class extroverts). This is unrepresentative sampling, so the generalisation of my findings is limited.

7.6.6. Technical management

There were a number of practical issues in running my experimental design. There was a lot of technical management required in use of equipment such as synchronisation of recordings, and early mistakes meant I had to discard some samples and recruit more subjects to make up numbers. Some technologies that would have made control much easier (such as remotely controlled swivel cameras) were too expensive for me to obtain. At the time of running my studies, the size of video files was an issue for the capabilities of the PCs at my disposal, causing some system crashes and slow running and making editing and reviewing files time consuming. At the outset, this was quite a new field so creating links and a community of interest and support has been very important.

7.6.7. Masked expressions

I was struck by a significant number of subjects whose emotion related behaviours were at least partly concealed by their posture or accessories, (as described in Chapter 5) such as with faces hidden by baseball caps or hoodies and expressions masked by chewing gum. A human controller was largely able to cope and compensate for these factors, but a machine relying on detection of particular facial features may well have been confounded and would need significant repair of missing data.

7.7. Coding Method and Validation

Considerable work went into developing a novel method for coding user emotional expression during interactive experiences. When considering analysis I recognised shortcomings of existing coding approaches and hypothesised that a technique derived from those used in interaction analysis and spoken language transcription might be beneficial in this context. The method involves identifying affectemes (which are analogous to phonemes in language transcription). Finer grained analysis can also be performed to identify variations within any one affecteme (allafects which are analogous to allophones in language transcription).

This method enabled me to identify and code the behaviours I was interested in, in context, during HCI and to consider their components. Although observational schemes have limitations, this method has provided a structured way to assess observational data and to objectify and quantify it.

Experience with using this coding scheme suggests that it is straightforward to apply, is quicker than alternatives such as FACS and is easy to learn. I also found that it provided useful insights into the patterns of emotional expressions displayed at the user interface. Preliminary evidence on the reliability and validity of the method is also encouraging. Further work is needed to explore these issues in more detail.

My method succeeded in showing differences between affectemes. By adding further levels of coding, more sensitivity could be achieved. For example it would be useful to use this method to identify segments for detailed coding using more established coding schemes, such as FACS.

I went to considerable lengths to validate my coding scheme. I noticed some interesting variation in coding, such as recognition of the exact time of onset and offset and the numbers of episodes identified. Future analysis needs to consider sequential and temporal patterns in the presentation of affecteme components from various modes and the onset, peak and offset of actions, and the likely transitions between levels of intensity and valence.

It is tempting to ascribe underlying emotion to affectemes. However the same emotional expression may have different underlying meaning. It might be useful to adopt a structured framework to try to appraise meaning, in context, and this is another direction for future work. Researchers are now exploring the use of the Appraisal Analysis Framework (Martin & White, 2005) (initially designed for analysis of words) for analysis of images. It might be possible to adapt this for analysis of non-verbal behaviours.

This method provides an approach to coding affective and expressive behaviours during interactions with computers. While the basis for the technique is not new, drawing as it does on well-established techniques from interaction analysis, this type of approach may be unfamiliar to some HCI practitioners and its application in this domain is novel. In addition I have provided a new vocabulary for discussing units of emotion related, affective or communicative behaviour, drawing upon the well-known distinctions made in spoken language transcription. I hope that providing this vocabulary will facilitate further research in this area by providing researchers with a common language with which to describe their findings.

I have begun to identify some of the affectemes shown during human-computer interaction, (and which may, or may not be specific to my puzzle solving task), e.g. the fed-up ‘chin dump’, the anxious ‘peer’ and the turn taking ‘postural twirl’. Work on identifying affectemes is in its very early stages but it may eventually be possible to define a dictionary of meaningful units of affective behaviour, or HCI affectemes, relevant to different domains. By describing my work here and presenting a vocabulary for such analysis, I hope to encourage other researchers to contribute to building up such knowledge. There is potential for a database of video clips showing dynamic representations of all the different ways an emotion can be shown, or of all the different meanings that one emotion related behaviour has.

In my validation exercise I noticed that male coders tended to identify fewer episodes than females. This is in line with gender differences reported elsewhere, for example that females are better than males at recognising emotional expressions and that females recognise emotions equally well from male and female faces (e.g. Thayer & Johnsen, 2000; Penton-Voak, Allen, Morrison, Gralewski & Campbell, 2007; Hofmann, Suvak & Litz, 2006). I dealt with this by using female coders. In future work I will explore varied abilities of coders and in particular gender related issues.

The explosion of brain research using new methods such as fMRI and MEG has helped us to understand how the brains of differently-able individuals vary in function, for example the areas that are activated in individuals with synaesthesia (e.g. Nunn, Gregory & Brammer, 2002). Meanwhile the rise of the internet has allowed individuals with different difficulties to find others with similar problems and so to identify their problems (e.g. propopagnosia or 'face blindness' see Burman, 2002). Conditions once thought very rare, such as face blindness are now thought to be more of a continuum. Those who are coding during research should probably test their sensitivity and skills in recognising faces, expressions and affect and report that along with their coding results. It might be useful to identify and develop and use some standard assessments for researchers carrying out coding to use and report on themselves in this way.

Establishing reliability of coding is complex and there is no ideal recommended method that gives credit for close rating scores, when using Likert type scales. Several refinements of procedures were needed in order to bring agreement up to an acceptable level. Whilst I found the Self Assessment Manikin was a very useful tool, improved reliability may be achieved by devising a less sensitive scale, collapsed to fewer points. I found differences between pairs of raters dependent upon gender of the raters and participant samples. In the early stages of agreement rating, three of my raters were female and two male and the subject sample were 3 male and 3 females. There was higher agreement between pairs of raters of the same gender, particularly when they were rating the opposite gender. Female agreement was greater than agreement between males. This is consistent with the literature on gender differences in the perception of affect. For example from an early age and across many cultures, females have greater ability than males to perceive facial expressions of emotion (Elfenbein, Marsh & Ambady, 2002). Males have particular inability to recognise females' negative displays and ability to recognise facial expression may also be affected by menstrual and testosterone cycles (Goos & Silverman, 2002). This might have important implications for standardised collection of observational data on affect, and larger studies are needed to investigate this.

7.8. Design Implications for Affective Systems

My study provides a reminder of the importance of not over complicating systems. If a 'help' button works well, and users like it, then there is no point in trying to offer another design element to serve the same function that is more complicated and requires more cognitive effort. Participatory, iterative design methods, starting with low fidelity prototyping should help identify and eliminate issues like this at an early stage.

Personalised design is another issue. I was very strongly aware of individual differences and requirements of participants and exploring personalisation of systems may be productive, both using different affective 'skins' for an application and designing systems so that users can vary their use and interpretation of the system (e.g. following Brinkman & Fine, 2005).

I reflected that my 'help' button, in some ways acted as an affective button. Users could hit it to show when they wanted a clue. My exact understanding of their emotion was not really required for this interaction. If they were excited and happy and wanting to race through the game, or miserable and frustrated because they were stuck on a clue, it was sufficient for this application to know that for some reason they wanted a clue. Designing simple ways (or the very simplest way) for them to communicate might be a way forward. Future work could explore the use of 'Affecticons', such as emotion buttons or other interactive devices that allow communication of emotion related states or desires.

7.8.1. Terminology

It was clear from my work that terminology is a real issue in this domain. Even after years of work, eminent researchers continue to struggle to reach consensus on definitions and the use of terminologies and taxonomies. What ordinary people and participants understand by 'emotional' terms and their attitudes to affective systems is another thing again. It is difficult to explore issues with users when there may not be a shared understanding of the language in use. My work led to an affective computing installation in the futures gallery of a science museum. I carried out a pilot survey of visitors and tried to assess their views on affective computing. During this exercise I found that people better understood terms like ‘feelings’ and ‘body language’. I used an adapted scale originally designed to assess attitudes to technology in the early days of computers. This pilot study lacked sufficient numbers for statistical reliability and remains unpublished but I feel there is potential for a larger scale survey on these lines to explore the factors about emotional machines that are of concern, or interest to ordinary people and so give more insights for design of future systems.

I was also struck by individual differences, in use of words and wondered what underlying shades of meanings they might reflect. Baron-Cohen (2003) uses an analogy with colour perception. It is very difficult to establish if people using a colour name are perceiving and experiencing the same phenomena, and that is also true for emotions.

7.8.2. Ethical Issues

There are a number of ethical issues arising around potential for good and potential for harm. Manipulation of emotions has always taken place but technological enhancement of this is new. There could be uses of ‘nudge’ techniques (Thaler & Sunstein, 2009) for altruistic reasons, such as to promote healthy behaviours – or for nefarious reasons, such as promoting crime or terrorism. Circumstances could arise where humans try to deceive the computer or vice versa. There are significant issues around trust and privacy that require exploration. Human reaction might also vary according to the success of the technology in detecting emotion accurately. There are different levels of human display, such as in mixed messages, for example where people say one thing but show from their tone of voice that they mean another. It may be detrimental if emotion is detected wrongly, or indeed, if too accurately.

Technological response to emotion, and the intention of the response, may also affect interaction. For example if the response is written, spoken, or shown by facial expression of an avatar, and intended to enhance or decrease the human emotions shown.

7.8.3. Tools and Methods

I have noted and welcomed an explosion of new tools and methods in this domain, but I have identified several areas where more work is needed. Coding is very lengthy and costly and likely to be required for some time to come, as automated systems are still in their infancy. Coding is usually done through a series of passes and iterations. A tool that allows any improvements on this would be most useful. I have identified a number of useful elements in existing tools and I hope in future work to incorporate them into one tool with an added haptic element. There is also a requirement for sequential statistical analysis methods to add quantitative rigour to analysis of qualitative data. I have explored some available methods and hope to do further work in this area in the future.

7.9. Summary

My post hoc experiment intended to validate decisions made by the Wizard, was not successful but did teach me many lessons, and emphasised the need to consider the complexity of emotional expressions that are often unconscious and their interplay between participants and researchers. There are particular implications about coding issues and the need to produce reliable and valid coding results and to develop procedures and tools for doing that.

Numerous issues of interest have arisen from my exploration of emotion in computing, including how very complex emotional interaction is to study. In our day-to-day activities we constantly, and largely unconsciously, monitor constant streams of communication that might or might not have emotional content and intent. We constantly interpret and infer the emotional states and underlying feelings and intentions of others. But it may be that bringing this into consciousness interferes with the process of emotional communication. Training in emotional behaviours is difficult and the interplay of participants and researchers may contaminate behaviours under study. Study design has to be very carefully thought through and piloted. WOZ methods are helpful but baseline behaviours and skills of both participants and researchers need to be established, particularly when coding behaviours.

There are many issues that will impact on design of affective systems and we need to standardise the terminology we use and remain aware of the wider public understanding of the wording we decide to use. As the use of various terms becomes accepted ‘jargon’ in affective systems and emotion research, we cannot be sure that the general public will understand the same meaning of common words like ‘emotion’ or ‘affect’. There are many ethical implications of developing systems that could be very persuasive and used for either good or evil purposes. We need more tools to support this domain of work and should be aware that many experiments carried out in universities using student participants may not translate to real world settings. A high level of technical management of studies is needed to overcome real world challenges posed by contexts ‘in the wild’ and management of participants. In such settings participants often behave in non standard and unexpected ways and might mask the very emotional expressions we set out to study and iteratively change and adapt their behaviours even during the course of the experiment.

In Chapter Eight I will draw some conclusions arising from the work, describe the contributions it makes and consider opportunities for follow on work.

8. Conclusions and Contribution

8.1. Introduction

The introductory chapters of this thesis showed how an area of interest was identified, how the literature was reviewed and how the research questions were set. Chapter Three described how methodologies were explored and decided on and Chapter Four showed how a study was implemented using an experimental design. Chapter Five presented results from the study and Chapter Six described a follow on experiment that used a revised protocol. Chapter Seven discussed the lessons learnt and issues arising from the experimental work. Throughout the explorations of this area a number of salient issues claimed attention.

This chapter describes the benefits arising from my work and the contributions it has made to the field of affective computing. Section (8.2) describes the contribution made by findings from the experimental work; design of the affectemes coding scheme; and a number of wider issues including identification of some limitations of the work. A chart summarises the contributions. The following section (8.3) describes ideas for future follow-on work arising from this research. Finally (in 8.4) the conclusions are summarised.

8.2. Contributions

My work primarily contributes to the field by confirming suggestions that affective computing has the potential to improve human-computer interaction. Previously empirical support for this claim was limited due to the relative immaturity of the technology to enable emotion recognition. My work was the first to show statistically significant benefits in performance and satisfaction when using affective systems. The results are notable, as they support the importance of ongoing work to enable emotion recognition at the user interface. My development of the 'affectemes' coding scheme allowed me to explore the data in a useful way. I have identified wider issues such as the usefulness of WOZ methodology, and I have illustrated the sheer complexity of emotion related behaviours and communication, limitations, design requirements of systems, coding issues and ethical concerns that need to be addressed during research into affective computing.

8.2.1. Objectives Met

The general objectives of the research identified from my review of the literature in Chapter Two were:

1. To establish the extent to which people will naturally express emotions when they know they are interacting with an emotion-detecting computer.
2. To identify the conditions under which the application of emotion detection can lead to improvements in subjective and/or objective measures of system usability.
3. To establish if and to what extent participants are willing to adapt their behaviours during interaction with an apparently affective system.

During the experimental design in Chapter Four these general objectives were developed into more specific questions:

1. What is the effect of interacting with a system that acts affectively on expressions of emotion?
2. What is the effect of interacting with a system that acts affectively on adaptation of user behaviours?
3. What is the effect of a system that acts affectively on task performance of users?
4. What is the effect of a system that acts affectively on reported satisfaction of users?

And the specific questions resulted in five hypotheses:

H1. If the system acts affectively, there will be improved performance by the participant.

H2. If the system acts affectively there will be improved satisfaction on the part of the participant.

H3. If the system acts affectively, then participants' expressions of emotion will increase in valence.

H4. If the system acts affectively, then participants' expressions of emotion will increase in intensity.

H5. If the system acts affectively, then emotion related behaviours will change.

Using empirical methods and experimental design described in Chapter Four, I demonstrated and reported in Chapter Five that:

1. Task performance was different when using an affective system: if the system acted affectively, there was significantly improved performance by participants.
2. Subjective satisfaction was different when using an affective system: if the system acted affectively there was significantly improved satisfaction on the part of participants.
3. Subjective satisfaction was different when using an affective system as shown by self rated valence of participants' emotional experience (see 5.3.2): if the system acted affectively, then participants reported that their expressions of emotion were increased in valence. After the game, while completing the post experiment questionnaire, participants reported themselves as significantly happier when they had used a responsive system.
4. Arousal levels of participants' emotional experiences differed when using an affective system: if the system acted affectively, then participants reported that their expressions of emotions increased in intensity.
5. Participants' behaviours were either observed or felt by participants to be different when they were using an affective system: if the system acted affectively, then emotion related behaviours changed. Participants reported that they thought they showed emotion significantly more when they were told the system would respond to them, and most of all when they were told it would respond, but it did not do so. This is particularly interesting as it suggests that participants were willing to try and repair the interaction by intuitively increasing their efforts to show emotional expressions. Participants blinked significantly more if they were told the system would respond to them. Participants' behaviour was rated as more positive when the system responded to them, and most positively of all for the group who were also told the system would respond.
6. Behaviour was rated as more intense for participants who were told that the system was affective.

The results described here demonstrate that significant improvements in both task performance and users' subjective feelings can be achieved as a result of adapting an application on the basis of affect recognition at the multimedia interface.

Results show that users were both willing to adapt their emotion related behaviours and did in fact do so, for example by showing more emotional expressions when interacting with an affective system and increasing their expressions even more when the system they expect to be affective seems not to be working, as if they are trying to repair the communication.

Results show, that overall, participants reported they felt that affective recognition technologies should be developed, although a number expressed reservations about its desirability.

My evidence lends some support to the Media Equation principle (Reeves & Nass, 1996), that people treat computers just like they treat real people. The experimental application used was a traditional desktop game, with no obvious human characteristics, (for example with no avatar on the interface). Despite this, participants varied their responses to the computer system, just as they might be expected to vary their responses during human-human interaction. When two people are conversing and the conversation fails in some way, it is natural for one communication partner to support the communication of the other. For example a listener will nod their head or make affirmation noises or use affect bursts (such as 'hmmm') to encourage the person whose turn it is to speak. Or if something is said that is not clear, they will ask for clarification. If the repair is not successful people increase their efforts and take on more of the burden of the conversation. In the interaction with the experimental system it was particularly striking that participants put more effort into their emotional expression when the system failed to respond to them as they expected – rather as if they were using repair strategies in conversation.

8.2.2. 'Affectemes' Coding Scheme

In order to manage my data, I developed a novel coding scheme to annotate the observed episodes of affective behaviour. This borrowed from methods tried and tested in the social sciences and in analysis of verbal human communication. First, relevant episodes of affective behaviour (affectemes) are identified. The exact nature of the behaviours can depend on what is considered important for a system, so for example the coders could identify one or more specific behaviours, such as leakage of basic emotions (e.g. disgust or joy), or they could identify episodes of positive affect or episodes showing more complex affective states, such as boredom. Affectemes of interest can then be analysed in greater levels of detail, as required. I used three levels of analysis:

- verbal and non verbal behaviours;
- the mode of communication (e.g. whole body movement, gaze, facial etc.);
- the manner of communication (e.g positive or negative valence, arousal, etc.).

'Affectemes' is a structured coding scheme, but is very flexible, allowing researchers to use different levels of detail and to adapt the detail to their interests. Following Bakeman and Gottman's (1997) procedures, the scheme can be adapted for particular emotional behaviours. I have shown that this coding scheme has potential to be very useful in the analysis and evaluation of affective systems. Manual coding is a notoriously lengthy procedure. By using three levels of analysis it is possible to shorten the time required. The first pass is low in coding demand and therefore relatively quick to execute and it allows identification of episodes of particular interest. Subsequent passes are more detailed and lengthy but are concentrated on episodes of interest. By drilling down in this way the greatest detail is applied to the most critical parts of behaviour. Tools to support this kind of coding could speed the process.

8.2.3. Wider Issues

Using WOZ Methods. In my experiment I was able to successfully simulate (using a Wizard of Oz set up) the proposed capability of future systems to recognise affect from observed user behaviour. This study showed that the use of Wizard of Oz approach can be used effectively to simulate an affective system. In this case the affective nature of the system was manifested by recognition of displays of emotion-related communicative behaviours and a response to these by sending clues when participants appeared to desire them. This system was used successfully to explore user behaviour, preferences, satisfaction and usability gains when using systems that appear to recognise emotions. The WOZ methodology was simple and successful in that participants believed the application recognised and responded to their emotional expressions.

Complexity of emotion related behaviours and communication. I identified complex behaviour patterns that involved simultaneous, multiple modes to convey messages, for example facial expressions at the same time as postural movements and vocal sounds. A clear overview of the context was essential to make sense of changing patterns of emotion related behaviour. This has important implications for affective computing. System developers will need to consider many different ways in which emotions might be displayed and the subtle nature of emotion related behaviours and the social and cultural rules governing emotional displays. The same gesture or facial expression might carry completely different meanings depending on its context and conjunction with other signals – for example tears could be tears of joy or tears of rage or a smile might be a signal of a polite greeting rather than a sign of happiness. Or an expression made by one particular individual to another, might mean one thing to a stranger and something quite different to someone who knows that person and their habitual expressions well. A fleeting expression might be 'leakage' of an underlying emotion but it might be more polite for a system to ignore it than to respond to it. A smile might seem polite in one culture but rude in another.

Contribution to the terminology and models debate My work highlights some of the difficulties in the uses of terminology within the field of emotion recognition in computing and contributes to the ongoing debate about definitions and their uses and their fit to existing models of emotion. Work that illustrates the multimodal and interactive nature of emotional interaction has helped to drive the field from early reliance on basic emotions (e.g. Ekman, 1999) to interactional models (e.g. Boehner, DePaula et al, 2005). Work that illustrates the importance of users and the complexity of their needs and behaviours has driven a move from the design of quite prescriptive systems that sought to assume how the user felt and respond to those systems (e.g. Riseberg, Klein et al 1997) to much more participatory design, and affordance of collaborative interpretation, with users, to interpret signals detected by the system (e.g. Leahu, Schwenk et al, 2008; Picard, 2009). In particular, this work brings considerations from human communication science and linguistics, to help understanding of the complexity of emotion shown in human communication and in human-computer interaction. It leads towards an improved conception and consensus about terms and concepts in the domain of human computer interaction.

Limitations I identified some of the difficulties of involving users in discussion about affective systems and the problematic nature of making affective behaviours intentional and bringing them into the conscious domain. Participants were not always sure what was meant by the word ‘emotion’. Some participants interpreted this as inner feelings. Most participants were not usually aware of their own use of emotional expressions. Some said that they intentionally tried not to show any expression – although observers could still identify changes in expressions. Training participants to use particular emotional expressions proved difficult. As participants lacked experience in using affective systems and lacked awareness about the ethical and other ramifications that might arise from such development, it was very difficult for them to respond meaningfully to some of the questions posed – such as those relating to whether they would welcome the development of affective systems.

Requirements for design of affective systems Involving users in design enables them to understand their preferences and gives them choice and control. As it is difficult to discuss systems because of the lack of common understanding and terms in this field, it is essential to involve users in participatory design and to let them try out different instantiations so that they can make informed choices about what they want, and so that developers can make informed decisions about what to develop.

Coding issues As accurate automatic coding of the multiple modes of communication that relate to emotional expressions and their context is a long way off, there is going to be a continuing need for manual coding for some time to come and so also a need for schemes and tools to support it. For coding to be reliable it is necessary to establish inter coder agreement. This is very complex when multiple aspects of communication are to be coded. In this study, the best agreement was found to be between female coders when rating male participants. There are issues about the variation between coders that require consideration and further analysis. As described in Chapter 2, there are many individual differences relating to experience of emotions and their expression and comprehension. People vary in their perceptual abilities and sensitivity to affect, and there are well established differences for example relating to gender and age. Coders prove an exception to this so it is important to develop tools to screen coders so that any biases are taken into account. There is an important need to screen for baseline levels and checks on sensitivity and skill of coders and reliability and validity of their coding.

Ethical concerns Affective systems can improve performance and satisfaction and so have potential to enhance human computer interaction, but there is also potential to cause harm. Developers might be able to design complex recognition algorithms that result in very convincing systems that are intended to support users, for example robotic companions for the elderly or infirm intended to make people feel happier and more secure. But the same kinds of systems might be used for psychological manipulation of populations by dictators. Just as guns can be used for protection or for mass murder, affective systems might be used for good or evil. Developers and designers should consider the potential uses of their work.

8.2.4. Issues for Future Work

I identified a large number of issues for future work (detailed in 8.3) that involve further trials of the work described here, re-use of data, development of new methods and tools, and studies to explore particular aspects of affective computing.

8.2.5. Contributions Chart

This diagram shows contributions from this work to the academic community.

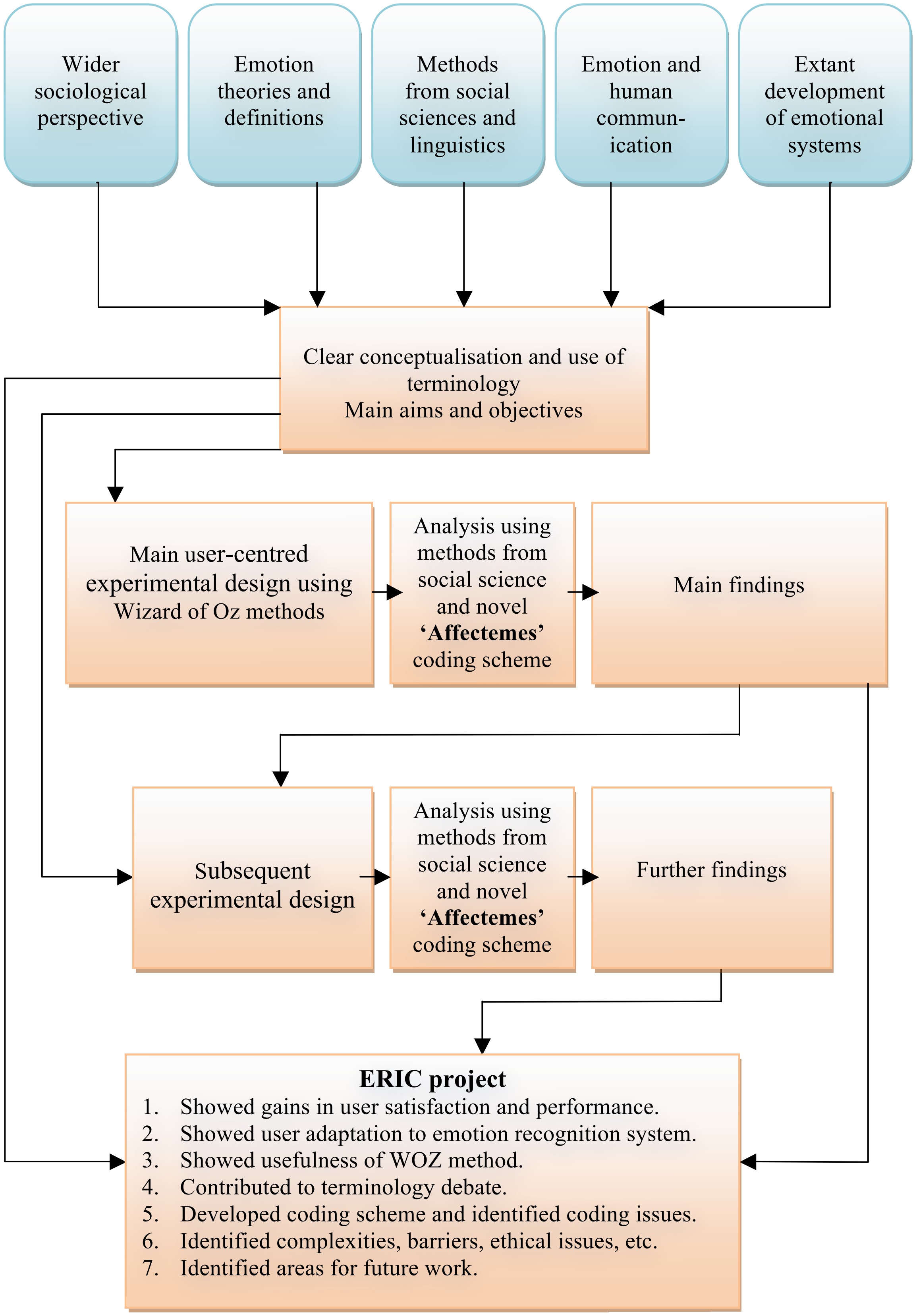


Figure 18. Contributions Chart

8.3. Future Work

I have identified a large number of areas for future work:

1. Validation of responses made in WOZ scenario. More work is needed to understand and validate the decisions made during the experimental work by the Wizard. The responses, that were made intuitively, seem to have been well received by participants, but they were made in response to a wide range of cues from the participants and it is not clear if they are replicable. An attempt to train participants to make certain cues and for the Wizard to respond only to those cues in a set way was not successful. More analysis is required to understand both the cues picked up by the Wizard and the responses made to those cues. If this could be achieved it would be useful for researchers developing algorithms for emotion recognition, along the lines of Natural Language Processing (NLP) approaches.
2. Need for work that considers multiple modes of communication. Affective systems research emphasis has so far concentrated on non verbal communication of emotions (including in my own work). Spoken and written language and how they interact with non verbal signs has so far been neglected in this research.
3. Creating a Library of affectemes and their variants. Any one emotional expression can be expressed in a large number of ways. This might involve different modes (e.g. tone of voice, facial expression, touch etc) or different timing. The same expression might mean different things in different settings or contexts or when communicating with different people. Despite this it would be very useful to start to build a library of expressions, annotated for context and intended meaning. This would be a useful resource for further research. While it would probably never be possible to state with certainty that a particular expression will always mean a particular thing, it might be possible to give a degree of certainly or a list of possible meanings.
4. Contagion issues. In the experimental work, it was noted that the participants behaved differently in the first minute after the researcher left the room. It took that long for them to switch from smiling and interacting with the researcher and to settle to the task. Emotional contagion might also carry over from the

emotional content of cues used. This requires further investigation. For example sequential analysis of behaviours in the context of their cues might show an increased likelihood of carry-over of positive or negative emotional behaviours in the segments following an emotional cue.

5. How studying affect and subsequent feelings of researchers and/or subjects might confound the affect itself. The long-term mood of the participant is very likely to alter their use of emotional expressions during the task and mood is very difficult to assess. Similarly the mood of the researcher might alter the interaction. A bad tempered or tired Wizard might provide cues differently, and coders might also vary in their mood, which might impact on coding decisions. This all needs more research and we need more methods to assess both long term mood and shorter term feelings.
6. Issues around profiling coders – there is variation in affect coding practice and we need to understand what factors impact on the coding decisions made and subsequent reporting of behaviours. We should profile – and publish information about the profiles – of those doing the coding. There may be long term differences between coders, e.g. some coders might be more susceptible to affect, or might have less skill in recognition of elements such as facial expression recognition. There may be shorter term differences between coders, or for one coder at different times, for example due to mood changes or fatigue on behalf of the coder or due to emotional contagion taking place. Variations in sensitivity may occur over time or in different conditions – for example an initial 'gut' feeling about meaning of a behavior might be more or less accurate than analysis of component parts of a complex behaviour. Until we have answers to such issues – or at least report them in a systematic way, along with our results… affect coding cannot be considered a highly reliable process.
7. Need for iterative user centred participatory design methods. As affective systems are a relatively new concept, and truly functional systems are in their infancy, we simply do not know how participants will feel about using them. In the validation experiment, participants did not find it easy or desirable to communicate via use of a specified emotional expression. Participatory design with the users might have identified this at an earlier stage. If costly mistakes

and undesirable systems are to be avoided, then iterative, user centred participatory design is an important route to experiment with.

8. Personalisation of affective systems (e.g. personalising for motivation). As individual use of emotional expression is very varied, then personalized systems might be one approach. In motivational systems personalization has been found to be very important (e.g. see Oulasvirta & Blom, 2008; Halko & Kientz, 2009) and this would be an interesting approach to explore with affective systems.
9. Training emotional behaviours. The attempt to train participants to use specified emotional behaviours did not prove successful, and one element of the problem might be concerned with training of emotional behaviours. There is a large literature on training social emotional behaviours in children (e.g.Rinaldi, 1995; Spence, 2003) and it would be interesting to explore methods for systematic training in the context of affective systems.
10. Designing affecticons. An alternative to complicated sensors or assessment methods to judge emotional state, might be simply to use an 'affecticon' or icon of some sort that users could click to indicate their emotional state. This possibility could be explored with users.
11. Need for more coding tools – e.g. haptic coding tool. Coding of behaviours is very time consuming and laborious, requiring many passes through any video footage to code different behaviours and rate them. For example in the experiment described in this thesis, a number of passes were made such as one pass to identify a period of positive valence and another to rate the degree of valence. A two or three-dimensional haptic tool might enable coding of degree of behavior at the same time as dimension and this is an area that needs more exploration and where I hope to develop a 'Flowcoder' tool.
12. Issues around identifying episodes – onsets, offsets and peaks. In the data collected for this thesis it was noted that emotional behaviours varied in timing of onset, offset and peak of behavior. This might be an interesting area to revisit and explore in more depth.
13. Sequential analysis of episodes. In a pilot exercise, GSEQ (Generalized Sequential Querier) (Bakeman & Quera, 1995) was used to explore patterns in

the affective data. It would be interesting to explore sequential patterns in more detail and breadth and using different methods and tools for this.

14. issues for sharing affective applications with various groups, when affect experience may vary with age, gender, culture etc. e.g. emotional experience and management are known to change with age, so it would be interesting to explore the requirements and performance of different intergenerational age groups with affective systems of different types.

15. Attitudes to emotion in computing. Attitudes to emotional machines are very difficult to establish for a number of reasons. Defining 'emotion' is not straightforward and as complex functional affective systems do not yet exist, the abstract concept of them proves to be quite a complex concept for researchers to communicate and for people to understand. In a small pilot survey, I used adapted versions of scales originally designed to assess attitudes to computers when those were a new concept. The numbers of participants were not large enough for valid statistical conclusions but factor analysis did suggest some trends. It would be interesting to revisit this and look at attitudes to different types of affective machines in different contexts.

16. Re use of my data. A wealth of data has been collected for the purposes of this thesis. Re-use is limited due to ethical constraints but it would be very interesting to look back at the data to explore specific questions in the light of new developments in the affective computing field. For example it would be interesting to compare coding of data using FACS, FEAT, Humaine definitions or other coding schemes. In time the UK Data Archive (2009) will undoubtedly have corpus data that can be used in multiple ways by different researchers.

17. More work using affectemes coding system. It would be interesting to explore using the affectemes system for different purposes. Using this coding method I have already re-visited the data looking at a small sample for evidence of instances of leakage etc. that might indicate that the participant is keeping secrets or telling lies.

18. Systemic Functional Linguistics and the Appraisal Analysis Framework. In linguistic anthropology, the Systemic Functional Linguistics approach is causing considerable interest and in particular the Appraisal Analysis Framework (see Martin & White, 2005). This approach emphasizes the

function of language in context. While mainly used for textual analysis it is now beginning to be used for visual analysis and there may be potential to adapt this framework to assess non-verbal communication in conjunction with spoken language. After a pilot attempt at using this method to identify where affect was indicated, I hope to explore this area further.

19. Work to consider the social construction of emotion – concept of distributed emotion. The concept of distributed cognition is well established (Hollan, Hutchins & Kirsh, 2000) and studies show how communication can be jointly constructed (e.g. Goodwin, 2004) and that emotion is interactional (Boehner, dePaula et al, 2005). From the observations made in this study, it seems that emotions are also constructed, not only between people but between people, artifacts such as the computer and the environment. It would be interesting to explore this concept of 'distributed emotion' further.

20. Economic evaluation of systems in terms of cost effectiveness, value added and well being. Affective systems offer promise in a number of areas including 'fun' areas such as gaming (which are a large business sector) and for 'serious' health and social care applications. Adding affective features might give a business advantage to systems but the exact benefits in economic terms for different aspects needs assessment.

21. Work with focus on particular elements of emotion e.g. motivation. Emotion is a very wide area, incorporating any number of emotional aspects. It would be interesting to explore specific aspects of emotional life. One aspect of current interest is design of motivating systems, e.g. for motivating behavior change and healthy lifestyles or using motivational interviewing in systems.

22. Ethical issues. There are a number of ethical issues arising from this work. Ensuring properly informed consent is complex when affective systems are involved. Affective systems have the potential to influence how people behave for better or worse. It is important that designers consider this. A system intended to influence people for good or in positive ways could equally be used to influence people for evil or in negative ways (See Kurzweil, 1999).

23. Multisensory combinations of stimuli. The work of neuroscientists is revolutionizing our understanding of perception and emotional functioning. For example Rowland and Stein (2007) have found that tiny stimuli from different

modes (e.g. visual and auditory), when co-located, can cause huge recruitment of perceptual signals. It may be that in the same way, combinations of very subtle synchronous emotional signals from different modes (both spoken and non-verbal) are recruited and better perceived than single mode, gross movements. This would be a very interesting area to explore.

8.4. Summary

In previous chapters I introduced the field of emotion recognition in computing and reviewed the literature connected with affective computing. I described my selection of methodologies and how I used them in an experimental study and the results yielded. I described some repeated trials using a variation in the protocol and the lessons I learnt from that.

In this chapter I have stated my contributions to the field, shown how I met my objectives and how I contribute to wider issues in this domain. I discuss issues arising from my work and limitations of it. I show how the issues raised led to many areas to explore in future work.

The appendices give further details about publications arising from this work, some sample publications that give more details about the experimental work and the 'affectemes' coding scheme, and provide samples of documentation used in the experimental work, including the information given to participants, pre and post experiment screening assessments and samples of data analysis.

The study of emotions in computing crosses the boundaries of many disciplines. These include physical and social sciences, such as philosophy, biology, psychology, neurology and applied linguistics as well as technological areas such as artificial intelligence, virtual reality, motion capture, visualisation and agent modelling. Contributions from all these different disciplines are important to consider. For example, recent advances in brain imaging have helped identify brain regions associates with discrete emotions. In future, remote mapping of brain activity might enable or enhance recognition of a particular emotional state. Development of Artificial Intelligence (AI) systems, Natural Language Processing (NLP) and neural networks will enable better recognition of the meaning of individual patterns of behaviour. Human-computer interaction, particularly involving affective behaviours with affectively responsive systems provides an interesting form of communication. New methods and tools and use of methods from other domains hold promise and linguistic methods such as Systemic Functional Linguistics (SFL) might enhance prediction and understanding of affective messages.

Systems are still a long way from human capability. Applications that support the user in a particular way and in a particular setting are likely to be far more successful than any unwise attempts at systems with wide ranging, but inappropriate or unrealisable human-like capabilities. As systems evolve, complex ethical issues and user preferences are arising and need careful consideration. We need to design further experiments in which we take account of the extremely complex issues and variables at play, new understandings of the neurological and cognitive basis of emotion and the fine line between what is emotion and what is communication. For effective and desirable affective system development, it is essential to keep the user focus at the heart of this work. We need to explore issues in relation to affective systems in order to gradually and incrementally improve the opportunities for human interaction and advancement with affective systems, so that they are of real benefit to people.

Products have been designed for emotional appeal throughout history and in recent times this has become a business and design science. Affective systems allow this design to go even further and become interactive (Cockton, 2009). Business advantages are perceived for systems with affective characteristics. For example the gaming industry is currently driving the development of technologies that recognise and respond to gestures, voice and facial expressions e.g. Kinect (Kinect website, Xbox, 2009). Affective systems might also promote health and well-being. With economic pressures on services delivered by people and with our growing emphasis on the ‘experience’ culture, it seems that affective machines are inevitable. Governments, businesses and others all want to ‘nudge’ our behaviours and affective machines may be perfect tools for this. It is very important that designers consider the many aspects of systems and their potential impact, for both good and bad.

This thesis explores some of the multiple aspects and shows how hard it is to consider them all. The technical challenges of designing accurate recognition systems for different modes such as vision, hearing, touch taste and smell are enormous, and designing systems for integrating information from all these modes is even more of a challenge. But even when this is achieved without error, getting recognition right is only the beginning. *Interpretation* of recognised sights, sounds, smells, tastes and touch all set in different time and contextual frameworks is far more complex. During interaction, iterative communicative responses will continually trigger changes in modes and meanings. We do not yet fully understand how people achieve smooth communicative emotional interactions in face-to-face interactions or how individual preferences and skills in emotional communication vary. Human to human emotional communication is difficult and can and often does fail. The impact of emotional communication breakdown can be devastating. We should not therefore, assume that design of affective systems is in any way easy or straightforward and we should remain mindful of the ethical issues surrounding their development and the benefits and pitfalls that might arise from our design and development efforts.

9. Appendices

Appendix A List of publications and dissemination arising from this work

Appendix B EPSRC assessment of final report on project

Appendix C Final draft version of CHI'05 paper: Axelrod, L., Hone, K. (2005) E-motional Advantage: Performance and Satisfaction Gains with Affective Computing. *Proceedings of CHI 2005* 'Technology, Community, Safety' Conference on Human Factors in Computer Systems, Portland, USA.

Appendix D Final draft version of BIT journal paper: Axelrod, L. Hone, K. (2006) Affectemes and allafects: a novel approach to coding emotional expression during interactive experiences. *Behaviour & Information Technology, special issue on Empirical studies of the User experience* Vol. 25, No. 2, 159 – 173.

Appendix E Advert, Game Design Details, Participant Instructions, and Consents, Pre and Post Experiment Scales and Question Sheets.

Appendix F Quantitative Analysis and summary of results.

Appendix G Qualitative Analysis

Appendix A. Publications and dissemination arising

Yr	Title	Author	Type	Published / disseminated
02	EPSRC Spotlight Issue 21 "How do you feel?"	Featuring Hone K	Magazine article	http://people.bath.ac.uk/pysabw/research/scell/scellsnewsline.pdf
03	BBC News item	Featuring Hone K	News item	http://news.bbc.co.uk/1/hi/technology/2861831.stm
03 On-wards	eMo Emotional robots exhibit. **	Featuring Sharkey N Hone K Axelrod L	Public display	Thinktank Museum, Futures Gallery, Birmingham. http://www.thinktank.ac/page.asp?section=239&sectionTitle=Find+the+Future
03	Letting your computer know how you feel	Featuring Hone K	Magazine article	Computer Weekly, June 24, 2003 http://www.computerweekly.com/Articles/2003/06/24/195402/Letting-your-computer-know-how-you-feel.htm
03	Emotional Access: User Preferences for Emotional Interaction in Computing	Axelrod L	Peer reviewed conference	British computer society (BCS HCI) Proceedings of 'Designing for Society' conference in HCI 2003 Volume 2 (doctoral consortium)
04	The Affective Connection: How and When Users Communicate Emotion	Axelrod L	Peer reviewed conference and poster	Proceedings of CHI 2004 'Connect' Conference on Human Factors in Computer Systems, Vienna, Austria (2004) (doctoral consortium)
04	Interacting with EMO	Axelrod L Hone K	Report	A report on visitors' attitudes to emotion in computing and their interactions with the affective computing exhibit for Millennium Thinktank Science Museum, Birmingham 2nd April 2004.

** EPSRC selected our project to feature in an exhibit displayed in the Futures Gallery at the ThinkTank Museum, Birmingham. The exhibit includes eMo, an emotional robot, built to illustrate our project by Prof Noel Starkey, EPSRC media adviser from Sheffield University. eMo is featured along with some interactive emotion games and a video designed to illustrate some of the issues around emotions in computing. The video features Noel Sharkey, Kate Hone and Lesley Axelrod. The eMo exhibit has led to much spin off publicity and inspired further work in the field e.g.
EPSRC Connect magazine Feb 2004 Issue 19; Real Robots magazine feature, 2004
http://www.epsrc.ac.uk/CMSWeb/Downloads/Publications/Connect/Connect19.pdf
Masters dissertation Anthony Bailey 2006 Synthetic Social Interactions with a robot using the basic personality model. http://www.dcs.shef.ac.uk/intranet/teaching/projects/archive/ug2006/pdf/u3ab.pdf
American Society for Engineering Education (ASEE) Egfi BLOG
http://students.egfi-k12.org/engineer-spotlight-noel-sharkey/
Radio 4 'Shoptalk' feature on AI. Tuesday 23 December 2003
http://www.aiai.ed.ac.uk/events/shoptalk2003/bbc-r4-shoptalk-20031223.html
Rudall B. H. (2004) New advances in biocybernetics. Kybernetes: The International Journal of Systems & Cybernetics, Volume 33, Number 7, 2004 , pp. 1084-1097

Yr	Title	Author	Type	Published / disseminated
04	Smoke and Mirrors: Gathering User Requirements for Emerging Affective Systems	Axelrod L Hone K	Peer reviewed conference	Proceedings of ITI 2004 26th International Conference on Information Technology Interfaces, Dubrovnik, Croatia.
04	Affecteme, Affectic, Affecticon: Measuring Affective Interaction with Standard and Affective Systems	Axelrod L Hone K	Peer reviewed conference	Proceedings of 18th British HCI Group Annual Conference 2004 'Design for Life' Conference, Leeds, UK
04	Emotion Recognition in Computing (ERIC)	Axelrod L Hone K	Poster	Brunel Post Graduate event
05	Affectemic analysis: User centred evaluation of multimodal interaction	Axelrod L Hone K	Position paper	Humaine Workshop. Work package 6, Emotion in Interaction, Paris, 10-11 March 2005 http://emotion-research.net/
05	E-motional Advantage: Performance and Satisfaction Gains with Affective Computing.	Axelrod L Hone K	Peer reviewed conference	Proceedings of CHI 2005 'Technology, Community, Safety' Conference on Human Factors in Computer Systems, Portland, USA (April 2005)
05	Emotion recognition in Computing (ERIC) project	Axelrod L	Invited seminar	Invited seminar for Interaction design group, Middlesex University April 19th 2005.
05	Identifying Affectemes: Transcribing Conversational Behaviour	Axelrod L Hone K	Peer reviewed conference	Proceedings of Symposium on Social Intelligence + Interaction in Animals, Robots and Agents April 2005 (Conversational Informatics) Society for Study of Artificial Intelligence and Simulation of Behaviour (AISB)
05	Development and Evaluation of an Empathic Tutoring Agent	Hone K Axelrod L Parekh B	Peer reviewed conference	Proceedings of SSAISB Symposium on Social Intelligence and Interaction in Animals, Robots and Agents (Virtual Social Agents) April 2005
05	Emotion Recognition in Computing	Axelrod L	Poster	BCS London Hopper: Women in Computing event

Yr	Title	Author	Type	Published / disseminated
05	Virtual Tea and Sympathy: User preferences and displays of empathy in HCI	Axelrod L Hone K	Position paper	BCS HCI Workshop: Empathy in Human Computer Interaction 19th British HCI Group 2005 Annual Conference: Design for life. Napier University, Edinburgh, UK, September
05	The Role of Emotion in Human-Computer Interaction	Peter C Axelrod L	Workshop organiser	19th British HCI Group Annual Conference, Napier University, Edinburgh Workshop: The role of emotion in HCI: design for emotional life
05	Beyond emotion: Using Appraisal Analysis to assess multimodal interaction	Axelrod L Hone K	Position paper	Proceedings of 19th British HCI Group Annual Conference Napier University, Edinburgh, UK, September 2005.
05	Uncharted passions: User displays of positive affect with an adaptive affective system	Axelrod L Hone K	Peer reviewed conference	Proceedings of First International Conference on Affective Computing and Intelligent Interaction, Beijing, China, October 05.
06	Engaging with Emotions – the Role of Emotion in HCI	Peter C Crane E Axelrod L Beale R	Edited proceedings	In Fields et al.(Eds.). Proceedings of the HCI 2006 Conference, London, volume 2, British Computer Society (2006). pp 270 - 272. ISSN 1470-5559
06	Affectemes and allafects: a novel approach to coding emotional expression during interactive experiences.	Axelrod L Hone K	Peer reviewed journal	Behaviour & Information Technology, special issue on Empirical studies of the User experience Vol. 25, No. 2, March-April 2006, 159 – 173
06	E-motional Behaviour: Multimodal emotional expressions in response to multimedia perception and design	Axelrod L Hone K	Book chapter	chapter in ‘Digital Multimedia Perception and Design’ Publisher: Idea Publishing Inc.. 2006. Eds. S. Chen, G. Ghinea.
06	The Role of Emotion in Human-Computer Interaction	Peter C Crane E Axelrod L	Workshop organiser	The 20th British HCI Group Annual Conference Workshop: The Role of Emotion in HCI Queen Mary, University of London

Yr	Title	Author	Type	Published / disseminated
07	Emotion recognition in Computing (ERIC) project	Axelrod L	Invited seminar	Invited seminar for UCLIC UCL interaction Centre University College, London, May 23rd 2007
07	Emotion in Human-Computer Interaction	Peter C Axelrod L	Workshop organiser	The 21st British HCI Group Annual Conference Workshop: The Role of Emotion in HCI Lancaster University
08	Emotion in Human-Computer Interaction: Designing for People	Peter C Axelrod L Fabri M Aigus H Crane B	Workshop organiser	The 22nd British HCI Group Annual Conference Workshop: The Role of Emotion in HCI John Moore University, Liverpool
08	Emotion in HCI: Joint Proceedings of the 2005, 2006, and 2007 International Workshops	Editors: Peter C Beale R Crane E Axelrod L Blythe G.	Edited proceedings	Fraunhofer IRB Verlag Stuttgart. First published: 2008 ISBN: 978-3-8167-7540-9 Available from http://www.emotion-in-hci.net/
08	Measuring affect: differentiating positive affect and politeness	Axelrod L Hone K	Position paper	CHI Workshop Measuring Affect Measuring Affect in HCI: Going Beyond the Individual
08	Back to Basics? Non verbal signals as indicators of secrets and lies in computer-mediated interaction.	Axelrod L Hone K	Position paper	CHI Workshop: Secrets and Lies
09	Putting the feel into coding: FlowCoder, a prototype haptic coding tool	Axelrod L Fitzpatrick G Mcallister G	Position paper	BCS HCI Workshop: Emotion in HCI: Real world challenges
09	Emotion in Human-Computer Interaction: Real World Challenges	Peter C Axelrod L Crane A Aigus H Balaam M Afzal S	Workshop organiser	The 23rd British HCI Group Annual Conference, Workshop: The Role of Emotion in HCI Churchill College, Cambridge
09	Under submission Emotion in HCI Real World Challenges	Peter C Axelrod L	Edited proceedings	Publication proceedings Workshop BCS HCI 'Emo4hci'

Appendix B. Copy EPSRC project report

EPSRC

Engineering and Physical Sciences Research Council

Research Support and Development Office
Brunel University
Uxbridge
Middlesex
UB8 3PH

Polaris House
North Star Avenue
Swindon, Wiltshire
United Kingdom, SN2 1ET
Telephone +44 (0) 1793 444000
Internet http://www.epsrc.ac.uk

Direct Line 01793
Direct Fax 01793
E-mail edward.whyte@epsrc.ac.uk
Grant Ref:GR/R81374/01

20/06/2006

Dear Research Support and Development Office:

Final Report: GR/R81374/01
Investigator: Dr Hone
Department: Information Systems Computing and Maths
Grant Title: Emotion Recognition for Interaction With Computers (ERIC): a Human Factors Evaluation

Thank you for the Final Report on GR/R81374/01. A summary of the assessment is shown below:

Criterion	Final Assessment	5-box Rating Scale
Research Quality	Tending to internationally Leading	Unsatisfactory to Internationally leading
Research Planning and Practice	Internationally leading	Unsatisfactory to Internationally leading
Potential Scientific Impact	Tending to internationally Leading	Unsatisfactory to Internationally leading
Quality of Training and Experience	Outstanding	Unsatisfactory to Outstanding
Communication of Research Outputs	Outstanding	Unsatisfactory to Outstanding
Potential Benefits to Society	Tending to Outstanding	Unsatisfactory to Outstanding
Cost Effectiveness	Outstanding	Unsatisfactory to Outstanding

Overall Assessment: Tending to Outstanding

Yours sincerely

Edward Whyte

Mr Edward Whyte
Programme Operations Directorate

22 JUN 2006

NX0116 v1.1

Appendix C. CHI'05 – draft version of paper

E-motional Advantage: Performance and Satisfaction Gains with Affective Computing

Lesley Axelrod **Kate Hone**
Brunel University
School of Information Systems and Computing, Uxbridge Campus, London UB8 3PH
UK
lesley.axelrod@brunel.ac.uk; kate.hone@brunel.ac.uk

ABSTRACT
Emotions are now recognized as complex human control systems, crucial to decision making, creativity, playing and learning. Affective technologies may offer improved interaction and commercial promise. In the past, research has focused on technical development work, leaving many questions about user preferences unanswered. For this user-centred study, 60 participants played a simple 'word ladder' game under different controlled conditions. Using 2 x 2 factorial design, and a Wizard of Oz scenario, half the participants interacted with a system that adapted on the basis of the user's emotional expression and half were told the system could react to their emotional expressions. We established that when using an apparently affective system, users perform significantly better and report themselves as feeling significantly happier. We also discuss behavioural responses to the different conditions. These results are relevant to the design of future affective systems.

Author Keywords

Affective computing; human-computer interaction; human factors; emotion recognition; Wizard-of-Oz.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Whilst the concept of emotion is poorly defined, the term 'affect' is now used to encompass this spectrum which includes feelings, moods, sentiments etc. Rosalind Picard distinguishes between inner emotional experiences and the outward emotional expressions that people may use to convey messages about their emotional states [5]. Like verbal language, emotional expressions relate to timing and context and follow social norms to convey information; they can be used to add emphasis to what is said, or for additional information, or without any verbal language. These paralinguistic emotional expressions are conveyed multi-modally by, for example, tone of voice, body movements, gesture and facial expression.

As computers become ubiquitous, and users' requirements include design for engagement, enjoyment, fun and playability, as well as usability, technologies may benefit from emotional design [4]. It has been proposed that future systems, such as e-learning and entertainment applications, are likely to benefit users by their ability to recognize human affective states and adapt accordingly. Systems that allow users to communicate their emotional states with others and systems that show or generate emotions are also under development [5].

Emphasis has been concentrated on the tremendous advancements in technologies, such as algorithms that enable computer systems to recognize human emotions and to show emotional responses by adaptation of the system pace or content or by avatars that show some emotions. Despite these advances, affective technologies do not yet approach human abilities to detect and respond to emotional expressions. Relatively few studies have therefore attempted to study user interactions with affective systems that respond to human emotion. The research described in this paper takes a user-centred perspective, examining some of the basic assumptions that underlie the development of affective computing applications. We were interested in determining user responses to affective computing interventions. Specifically we were concerned with whether such interventions could (a) improve performance and (b) improve user satisfaction with the system. We were also interested in the degree to which users would emphasize their emotional expressions in situations where they believed that the computer could respond to this. In order to investigate these issues an experiment was designed in which an affective computing application was simulated through the use of a Wizard of Oz (WOZ) set up (where a hidden human observer controls the computer's 'affective' response). Affective interventions were provided within the context of a game involving problem solving. The design of the experiment was such that in some conditions the system appeared to vary its response on the basis of recognized emotional expressions at the interface, for instance providing clues when the user displayed negative affect. We hypothesized that such interventions would lead to improved task performance and improved satisfaction. We also varied the conditions according to whether the user was explicitly told in advance that the system might react to their emotional expressions. This design allowed us to separately examine the effects due to the system response itself and those due to users' expectations. We predicted that those anticipating an affective application would show increased emotional expression during their interactions.

METHOD

Factorial Design

The experiment had a 2 × 2 between-subjects factorial design. The factors were:

1. ***Acted affective*** (with two levels; 'standard' vs. 'system appears affective'). This refers to whether the system appeared to act affectively. In the 'standard' condition clues and messages appeared only in response to the user clicking the 'help' button. In the 'system appears affective' condition, if the participant was observed via the one way mirror to use emotional expressions, the game was controlled so that appropriate clues or messages appeared to them.
2. ***Told affective*** (with two levels; 'expect standard system' vs. 'expect affective system'). This refers to whether the participant expected the system to act affectively. In the 'expect standard system' condition participants were told they would be testing the usability of a word game. In the 'expect affective system' condition they were told that they would be testing the usability of a word game on a prototype system, that might recognize or respond to some of their emotions.

There were therefore four experimental conditions in total, representing the factorial combination of the two factors.

The Game

Given words	Clues	Solution
HEAD		HEAD
	To listen	HEAR
	Animal that growls	BEAR
BEER		BEER

Figure 1. Structure of word-ladder game

A 'word-ladder' game was selected for the experiment as the task involves problem solving which is an area where affective interventions may be helpful. An initial word and a target word are given. Individuals have to attempt to transform the initial word into the target word by changing one letter on each of a number of given lines, or 'rungs', in response to clues, as seen in figure 1.

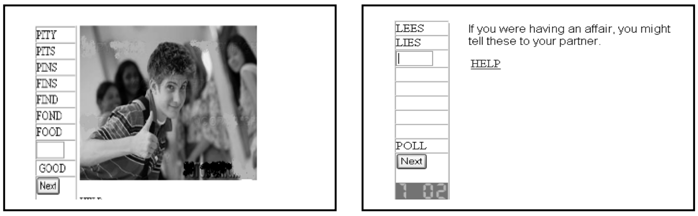


Figure 2. Screen shots of Word-ladder application

The game was designed to trigger episodes of user frustration and satisfaction. It involved the use of a slightly 'sticky' mouse, added time pressure and unexpected system interrupts. Clues during the game involved images and text likely to provoke some emotional responses. Figure 2 shows some screenshots from the application.

Participants

Sixty participants took part in the study (42 male and 18 female); fifteen were allocated to each experimental condition. All were currently living in the UK and most were graduate and undergraduate students. Participants were paid for their participation. The majority of participants were aged between 18 and 25. It was a pre-requisite that they be reasonably fluent in English. An attempt was made to match groups for cultural background and age.

Procedure

Before undertaking the word-ladder task, informed consent was sought and demographic data were collected. The degree of detail given about the purposes of the study was varied according to the experimental condition as described in the design. We were careful to ensure that although details were omitted in some conditions, no participants were given false information about the purposes of the study.
Participants attempted the word-ladder game, in either standard or affective mode depending on their allocation to one of the four experimental conditions. They were given 10 minutes to interact with the system.

In the 'acted affective' conditions WOZ techniques were used to simulate the capabilities of future emotion recognition technology. The experiment took place in a usability laboratory where two adjoining rooms, linked by a one way mirror were used, so the researcher could observe the participant without being seen. One room was reserved for the use of participants in the experiment, containing a standard desktop personal computer (PC) with a PC mounted video camera. The other room was reserved for the researcher and contained a PC with two screens attached. The PCs were linked by an area network so the researcher could view the participant's 'puzzle' screen and the context of the participant's interactive behavior at all times, as well as the researcher's own 'control' screen. When the system was intended to appear affective, the researcher could monitor and judge the affective state of the participant via the one way mirror, and adapt the game accordingly, so that clues and messages received by the participant were tailored to their emotional displays. All interactions were video-taped for later analysis and logs were collected of user keystrokes and mouse movements.

After undertaking the word-ladder task, participants were asked to rate their affective state on a nine point scale using the Self Assessment Manikin [1] which uses stylized figures to illustrate emotional state. They were also asked to complete a questionnaire which rated various aspects of their interaction, including whether they believed that they had shown emotion during the experiment.

As half the participants had not been initially informed about the role of affect in the experiment, there were some ethical concerns, so all subjects were fully debriefed at the end of their participation and apologies given for any failure to fully inform beforehand. It was made clear that they could withdraw their data from the trial, if desired.

Analysis

Task performance was measured by counting the number of completed 'rungs' on the ladder. Subjective satisfaction was measured from the ratings of emotional valence (from happy to sad) given on the Self Assessment Manikin. Behavioural responses were coded from the recorded observational data. Common behaviours such as smiling, frowning, resting chin on hand, shifting posture, grooming (e.g. adjusting hair) and blinking were identified, counted and rated for valence and intensity. The experiment's main research hypotheses were tested using analysis of variance (ANOVA).

RESULTS

Task Performance

ANOVA showed a main effect of system affective response on task performance ($F(1,56)=7.82$, $p<0.01$). Participants were able, on average, to complete significantly more rungs of the puzzle with the affective intervention.

Subjective Satisfaction

There was a main effect of affective response on user subjective ratings of their affective state after the interaction ($F(1,56)=10.25$, $p<0.005$). Participants reported themselves as significantly happier, on average, after interaction with the system that responded to their emotional expression.

Behaviour

There was a main effect of being told that the system was affective on participants' self ratings of the extent to which they had displayed their emotions ($F(1,56)=4.34$, $p<0.05$). Participants were more likely to say they showed their emotions when they had been told that the system was affective. There was also a significant interaction effect with whether the system had provided an affective response ($F(1,56)=4.34$, $p<0.05$). The highest self report of emotion was from those participants who had been told that the system might respond to their emotional expression, but where the system had not in fact been affective. This interaction effect is shown in the interaction plot in figure 3.

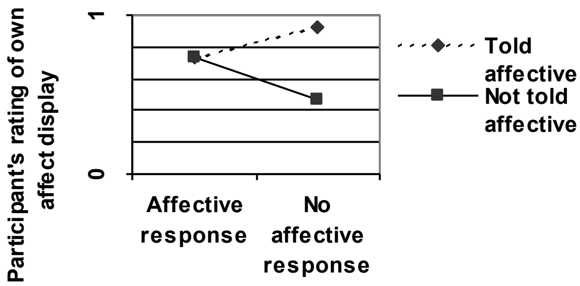


Figure 3. Interaction plot for self reported emotion display

Of the macro-level affective behaviours that have been coded, only blink rate was found to vary with experimental condition. There was thus a main effect of being told the system was affective on blink rate ($F(1,56)=4.57$, $p<0.05$). Blink rates were significantly higher when participants were told the system may respond to emotional expressions.

DISCUSSION

The results described here demonstrate that significant improvements in both task performance and users' subjective feelings can be achieved as a result of adapting an application on the basis of affect recognition. Participants performed significantly better and felt more positive when using the adaptive affective recognition version of the application This confirms suggestions that affective computing has the potential to improve human-computer interaction. Previously empirical support for this claim was limited due to the relative immaturity of the technology to enable emotion recognition. In this experiment we were able to simulate (using a Wizard of Oz set up) the proposed capability of future systems to recognize affect from observed user behaviour. The results are notable as they support the importance of ongoing work to enable emotion recognition at the user interface.

The research also considered how users' emotional expressions might vary when they use an affective system. It has been suggested that, in the short term at least, emotion recognition technology will need users to exaggerate or deliberately pose their emotional expressions [3]. We were therefore interested in whether telling a user that the system might respond to their emotions would lead to a greater degree of emotional expression at the interface. We were also interested in whether users might naturally adapt to systems which respond to emotion by displaying more expression, even when not specifically aware that the system might adapt.

It certainly appeared from the subjective data that participants believed that they were showing more emotion in the conditions where they had been told the system would respond to this. The highest self ratings of emotional expression were from those participants who were expecting an affective response, but where the system did not adapt. These results suggest that users of an affective system will exaggerate their emotional expression when they expect an affective response, and that this effect will be particularly pronounced where the affect recognition does not appear to be working.

Interestingly the self ratings for use of emotional expressions were as high for those who did not expect an emotional response as those who did, in the cases where the system did actually respond to emotion. This suggests that participants may adapt to affective systems by showing more emotional expression, even when not specifically expecting an affective response. Collectively the results from this subjective data are encouraging for the success of emotional recognition in the short term as they suggest that users will adopt behaviours that have the potential to aid the recognition process.

Despite the encouraging subjective results, we have found limited evidence so far for tangible differences in observable behaviour as a result of the experimental manipulations. It is clear from the video recording data that participants displayed a large number of affective signals at the interface. However, this appears to be true for all conditions, not just those in which the participants were expecting, or got, an affective response. Of the observational data that has so far been coded and analyzed, only one significant effect was observed, with blink rate found to vary according to whether participants expect the system to be able to respond to expressed emotion or not. Blink rate was significantly higher in conditions where participants were expecting an affective response. As blink rates have been linked to arousal and stress, these results suggest that the expectation of using an affective system may cause participants to feel more vulnerable or may add emotional load [2]. So far, analysis of the behavioural data has been restricted to macro-level behaviours that can be quickly and easily coded. In-depth coding of emotional expression is ongoing but is highly time consuming. It therefore remains to be seen whether experimental effects will be found on expression valence and intensity when coded at the micro-level. The lack of many clear influences of experimental condition on macro-level expressive behaviour, despite the significant subjective results, suggests that participants may to some extent misjudge the degree to which they display emotions when interacting with computer systems. There is some support for this contention from a minority of participants who stated categorically that, once told the system would be 'looking at them', they had decided to 'show no emotion'. In fact these participants showed many recognizable affective behaviours during interaction.

A possible limitation of this work lies in the use of the 'Wizard of Oz' technique in order to simulate affective responses. The experimental manipulation relies upon the experimenter making appropriate subjective judgments about when the system should make its affective interventions. Post hoc validation of the intervention protocols used is needed (and planned) on the basis of the detailed coding of emotional expressions. However, subjective data from experimental participants provides encouraging preliminary support for the validity of the experimental manipulation, with participants' post trial belief in the affective nature of the system varying by condition in the ways that one would predict.

Conclusions and Future Work

This paper has demonstrated empirically that usability benefits in both performance and satisfaction can arise from recognizing and responding to user affect at the interface. We have also reported on the behavioural responses of users of affective computing systems. These results have potential relevance to the designers of future affective systems. Detailed analysis of the observational data is ongoing and has the potential to further inform future design work. Further research is needed to establish if the benefits found here extend to other applications, platforms, and types of adaptation, and if so, to specify which ones. The Wizard of Oz technique used here appeared to work successfully to simulate the capabilities of future affective systems and could be used in the future to explore further questions within this domain.

ACKNOWLEDGMENTS

This work is supported by EPSRC grant ref: GR/R81374/01.

REFERENCES

1. Bradley, M.M., Lang P.J., Measuring Emotion: the Self-Assessment Manikin and the Semantic Differential. *J Behav Ther Exp Psychiatry* 25,1, 1994, 49-59.
2. Givens, D. B., *Nonverbal dictionary of gestures, signs and body language cues.* Spokane, Washington: Center for Nonverbal Studies Press, 2004. http://members.aol.com/nonverbal2/eyeblink.htm
3. Hayes-Roth, B., Ball, G., Lisetti, C., Picard, R. and Stern, A., Panel on affect and emotion in the user interface. *Proc Int. Conf.Intelligent User Interfaces*, 1998, 91-94.
4. Norman, D., *Emotional Design: Why We Love (Or Hate) Everyday Things*. Basic Books, 2004.
5. Picard, R.W., *Affective Computing*. Cambridge, MA: The MIT Press, 1997.

Appendix D. BIT journal - draft version of paper

Affectemes and allaffects: a novel approach to coding user emotional expression during interactive experiences

AXELROD, L. and HONE, K.S.
School of Information Systems, Computing and Mathematics
Brunel University, St John's Building, Kingston Lane
Uxbridge, Middlesex UB8 3PH

Lesley.Axelrod/Kate.hone@brunel.ac.uk
Tel: 01895 266009 Fax: 01895 251686

Abstract
The potential importance of human affect during human-computer interaction is becoming increasingly well recognised. However, measuring and analyzing affective behaviour is problematic. Physiological indicators reveal only some, sometimes ambiguous information. Video analysis and existing coding schemes are notoriously lengthy and complex, and examine only certain aspects of affect. This paper describes the development of a practical methodology to assess user affect, as displayed by emotional expressions. Interaction Analysis techniques were used to identify discrete affective messages 'affectemes' and their components. This paper explains the rationale for this approach and demonstrates how it can be applied in practice. Preliminary evidence for its efficacy and reliability is also presented.

Keywords: affect, human-computer interaction, facial expression

1. Introduction

As computers become ubiquitous, users' requirements include design for engagement, enjoyment, fun and playability, as well as usability. To achieve this, developers and usability professionals, committed to human centred and universal design, need to assess users' affective states during interaction. Some researchers, e.g. Scheirer, Fernadez, et al, (2002) propose using user affective state as a dependent variable in human-computer interaction (HCI) experiments. Others suggest that computers should be designed to automatically recognise user affect during interactions and use this information to react to user affect, with the hope that this may lead to more satisfying user experiences and improve the quality of the subsequent human-computer interaction (e.g. see Picard, 1997). Core to such developments are methods for measuring and coding affect at the user interface. We believe that within the domain of human-computer interaction it will be more fruitful to concentrate on the identification of communicative episodes, rather than necessarily seek examples of basic underlying emotion. Ward and Marsden (in press) present a similar argument. This paper reviews current approaches to this problem and proposes a novel method for coding user emotional expression during interactions with computers.

The paper proceeds as follows. We begin by explaining the current interest in measuring user affect during HCI, explaining why such information may be beneficial to system designers. We then go on to review current approaches to the measurement of affect, including physiological, subjective and observational approaches, drawing upon relevant theories of emotion where appropriate. Limitations of these methods are highlighted and we subsequently demonstrate how the method proposed in this paper was developed to facilitate analysis of affective data from an ongoing project. We describe the method itself and use examples from our project to illustrate the approach. Finally we present data supporting the utility of the suggested approach, present our conclusions and discuss future work that is needed in this area.

2. Background: The relevance of our emotional life to HCI

Emotions, once considered illogical, are now recognised to play an important part in decision making, in order to promote adaptive choices and aid survival, resulting in changes to individuals, to personal relationships, to organisations and to cultures. In 'Funology' Blyth, Monk et al (2003) bring together research that shows how users' emotions are critical in diverse domains such as e-learning, e-commerce and gaming. The importance of emotions increases as computing reaches beyond the desktop and office and into the realms of everyday life of individuals and communities, with ambient, ubiquitous, mobile, immersive and multimodal applications. Computers are no longer just data crunchers – we used them first for work – but now for education, fun, entertainment and communication – all important human functions. Hochschild (1983) describes the commercialisation of human feelings and Bunting, (2004) describes changes in organisational practices which require emotional labour and emotional management in workplaces in areas such as customer management. Organisations are interested in extending this to their computer mediated transactions. Fang and Salvendy (2001) predict the future thrust of research for e-business and HCI tools that will aid design of jobs and customer relationship management in the next decades. Tools to recognise and design for affect will certainly be essential for this. Given the emerging recognition of the importance of human affect in HCI it is relevant to include the measurement of emotion among the variables when evaluating interactive experiences.

3. The Measurement of Affect

3.1 Introduction
Lack of definition
A necessary first step towards measurement of affect is the initial definition of the construct itself. Many theories and models contribute taxonomies, or define basic emotions, dimensional models or more complex artificial intelligence process models, as reviewed by Mayne and Bonanno (2001). Emotion theorists such as Minsky (2004) now suggest emotions are complex neuro-physiological emotion systems with visceral, behavioural and reflective levels operating on biological, neurological and psychological systems. These systems help us organise and regulate other systems such as cognition, memory and problem-solving, critical to our everyday lives. Inter-disciplinary links are contributing to knowledge about emotions. Despite the additions to the knowledge base about emotions, terms such as 'emotion' 'mood' and 'affect' and thousands of descriptive word labels and metaphors are used inconsistently. They can refer to any or all of a spectrum of emotion related concepts and states, from personality types or individual traits to pervasive moods, fleeting feelings or experiences like pain or hunger that can generate emotions.

Experience vs expressions
In defining affect, a core distinction can be drawn between affect as an internal state, and its related manifestations as (a) the subjective experience of emotion and (b) the external display of emotion (such as through facial expression). Some measurement

problems arise because there are not always clear cut relationships between these three different ways of conceptualising affect. For example Argyle (1988) found it is common for people to display emotions which are at odds with their subjective experience of emotion. There are also suggestions that people may not always correctly evaluate their own internal state, leading to problems. For example, Schachter and Singer (1962) and Dutton and Aron (1974) found feelings of arousal may be misattributed as anxiety or attraction with consequent effects on behaviour. Assessing the underlying emotional experience is therefore extremely problematic. The subjective inner experience of emotion may not be fully reflected by the emotional expression a person displays. Even when two emotional expressions are identical, the subjective experience of emotion may differ between individuals or for any one individual over time, dependant on other factors such as context. Perception of the inner emotional state may vary between individuals. Even for perception of phenomena that can be exactly measured and classified, such as the colour spectrum or sound waves, we cannot be sure that any two people experience a stimulus as exactly the same. So is one person's experience of happiness the same as another's? Or is one person's experience of happiness more like another's experience of joy? Dimensional theories of emotion strive to overcome this difficulty of using word labels, by measuring emotions on different dimensions. Cornelius (1996) describes how traditionally separate schools of thought on theories of emotion have come closer together, so that many theorists agree that valence (whether the emotion is positive or negative) and arousal (how intense the emotion is) are interesting dimensions of emotion.

Our terminology

Language is dynamic and meaning of words change over time. In the domain of HCI Don Norman (2004) uses 'emotion' to refer to underlying levels and 'affect' to describe the reflective level. Using 'affect' as a broader, all encompassing term, Picard (1997) suggests that we can still usefully refer to, and examine, aspects of affect, even without complete agreement as to all its details. She uses the term 'emotional expressions' to describe what is revealed to others by affective displays of body language etc., and to differentiate outward expressions of emotion from the internal emotional experiences or states of individuals, which may or may not correlate to their emotional expressions. This pragmatic approach to the definition of affect is the one we shall take in the current paper. The focus of the current paper concerns assessing the external manifestations of affect, or what Picard (1997) describes as 'emotional expressions'.

3.2 Human Emotional Expressions

Human communication is dynamic, complex and rule based. We use, our background knowledge of individuals and their differences as well as different multisensory modes (Argyle, 1988):

- Appearance - (Occulesics) such as fashion, media products, architecture;
- Movement - (Kinesics) whole body movement, posture, tension, gesture, facial expression;
- Voice - (Vocalics or paralanguage) speech content, non speech sounds, choice of words, tone of voice, pitch, intonation contour, rate of speech;
- Touch - (Haptics);
- Smell - (Olfactics);
- Space - (spatial relationships or proxemics) territoriality, personal space;
- Time - (Chronemics) context, contagion.

There are different rules about use of emotional expressions for narrators and listeners, and variables such as personality, age, sex, culture, etc. It is a frequent criticism that much research on recognition of emotional expression is based on participants' observation of static images, rather than the dynamic interplay that arises in real communication situations, which Harwood, Hall et al (1999) found improved recognition of expressions. Emotional expressions often reveal information about underlying emotional states, attitudes and judgements, but they are also used to communicate ideas and facts, and can be used to control the flow of an interaction, for example in turntaking.

3.3 Existing Approaches to Identifying Emotional Expressions
Bio-physiological measures

Approaches to measuring affect as an internal state tend to concentrate on bio-physiological observations, using data such as heart rate, galvanic skin response and cardiac inter-beat interval. However the level of arousal shown by these measures does not map exactly to the subjective experiences of participants, and does not always give clear indication of valence, or exact emotional expression. For example Scheirer, Fernandez et al (2002) were able to discriminate highly frustrating interactions from non frustrating interactions on the basis of skin conductivity and blood volume pressure. However, Ward and Marsden (2003) report that physiological signals of this kind are not good for picking up more subtle HCI events. Despite the development of tools such as Kirsch's (1997) Sentic Mouse that detects affective valence from pressure, many such measures are quite intrusive and so may alter the participant's affective state. Functional magnetic resonance imaging (fMRI) is now providing fascinating data on brain activity during emotional processing, such as that in the Maddock, Garrett et al study (2003), or Mette, Posamentier et al (2003), but fMRI is currently restricted to laboratory settings which may not accurately reflect real life situations.

Probes and performance

There is a long standing history of assessing aspects of HCI through the use of subjective usability questionnaires or by measuring performance. However, typical notions of subjective satisfaction have not explicitly considered the concept of emotion in any detail. When affect is conceptualised in terms of subjective experience, measures can be obtained by asking someone about their feelings via direct probes, questionnaires or narrative means, either during interaction or retrospectively. However these very questions may disrupt the normal flow of affect, while semantic and perceptual differences between individuals mean that such measures may be imprecise. Retrospective recall of emotional states can be influenced by factors such as an individual's general conceptions of their emotions, or their coping styles. (Cutler, Larsen et al 1996, Scollon, Shigehiro et al, 2004). Performance measures may be useful to predict likely emotional states, for example to infer that a user is happy when a goal is met, but they do not measure actual user experience.

Recognising emotional expressions from observation

The recognition of emotional expressions from observation, has the potential to be particularly useful within the context of HCI because it is non intrusive and should therefore not disrupt ongoing interaction. In addition although emotional expressions may not always match internal affective state, they still typically convey useful information (i.e. they can reveal how the user wants to communicate their emotional state and other information to the outside world).

Automated recognition

Automated coding systems are progressing, which recognise and code specific aspects of emotional expressions such as facial movements or vocal pitch changes, for example, the Facial Expression Analysis Tool (FEAT) (Kaiser & Whurle 1992, 1994). Such tools are not yet available and reliable for every channel of emotional expression. Even when systems can reliably identify particular signs they will need contextual awareness to interpret intended meanings. Automatic analysis of spoken content and artificial intelligence systems should improve these tools in time.

Manual Coding

Numerous coding systems have been developed to help analyse specific aspects of emotional expressions. Some of these have been tried and tested for many years, such as the International Phonetic Alphabet for speech analysis, (Gimson & Cruttenden, 2001, IPA website,) and Labanotation for movement analysis (see Hutchinson, Hutchinson et al, 1987). Conversational and linguistic analysis schemes are also well tried and tested. Other well established schemes include Ekman's Facial Action Coding System (FACS) (Ekman and Rosenberg, 1998), Tone and Break indices (ToBi) (see ToBi website) and protocols for coding single modes, such as gaze analysis (ISLE, 2002). FACS is probably the most well known of the coding schemes for human emotional expressions. FACS is a technique for coding human facial expression, based on an anatomical analysis of facial movements. As the facial actions which form the basis of this method can be caused by the movement of more than one muscle, the basic units are called 'action units' (AUs).

However, Bakeman and Gottman et al (1995, 1997) report that to carry out detailed FACS coding from video takes between 8 and 100 times the duration of the video to fully analyze, making it impractical to attempt to code lengthy interactions for real world HCI practitioners.

4 The Proposed Method

This section describes the development of a novel means of coding multimodal emotional expression during interactive experiences. The method was developed during the course of analysing data from an ongoing study investigating the potential of future affective systems to improve the quality of human-computer interaction.

4.1 The Data Collection and Analysis Context

We are currently involved in a project investigating the potential for systems that recognise user affect to improve the quality of human-computer interaction. After a number of pilot observation exercises, the first experiment in this project was designed to assess the extent to which participants display emotion at the interface, either under normal conditions or when they believed the system could recognise and respond to their expressed emotions. A further experimental factor concerned whether the system actually altered its behaviour based on the emotional expressions observed (achieved via a Wizard of Oz manipulation) or did not. This paper does not aim to report the outcome of this experiment, but is focused on the approach developed in order to analyse the observational data and to enable a factor analysis (work which is ongoing). Preliminary results are reported elsewhere (Axelrod & Hone 2005).

The experimental task for this study was a simple on-screen word puzzle. This consisted of an on-screen word-ladder of simple four letter words, as in Figure 1. Each word could only differ in spelling from the previous word by one letter. Participants were given the first and last words of the ladder and offered a series of picture and text clues to help complete the rungs in-between. Time pressure was imposed by an instruction to play against the clock, with a limit of 5 minutes per game. Clues were designed to invoke some emotional reactions. Using WOZ techniques it was possible to tailor presentation of clues for individuals in selected experimental conditions, by preference for type of clue, (picture or text), and by degree of help or encouragement offered.

---insert Figure 1: Word-ladder puzzle task, about here---

The experiment was run as a 2x2 factorial design (illustrated in Table 1) with 15 participants in each of the four groups. In total we collected 60 ten minute samples of interaction for analysis. In order to test the experimental hypotheses these samples had to be coded for the presence of affective expressions, and for the valence and arousal levels of any such expressions.

During each trial users were video recorded for ten minutes each, using a JVC GR-D70 digital video camera and JVC miniDV tape. These were later rendered into MPG2 files at 29.97 frames per second. Videos and screen capture were edited together so that participants' behaviour could be viewed in the context of their interaction. All subsequent footage was reviewed and behaviours of interest were identified and discussed. After each trial demographic and subjective data were collected.

---Insert Table 1: Participant groups – 2X2 factorial design, about here---

4.2 Development of the New Coding Approach

As we reviewed the video footage collected from our project, it was apparent that participants were continually displaying emotional expressions. These ranged from very obvious events such as spoken words, phrases or vocalic affect bursts, (sighs, appreciative noises etc) to more subtle events indicated by body language of all types, including facial expressions. It was thus evident that multiple modes of communication were providing potentially relevant information. To assess multiple communication modes would require the use of multiple coding systems. Some of these coding systems also require frame by frame analysis of video sequences, which is extremely time consuming. Overall it was clear that to analyse the current data using the whole range of well know, tried and tested coding schemas (such as FACS) would impose a huge coding load on the project. Furthermore it was unclear what benefit such an approach would provide. The resulting data would be far more detailed than that required to evaluate the experimental hypothesis in the current work. In addition the existing techniques referred to above allow micro-analysis of different aspects of behaviour. Longer, meaningful pieces of behaviour for macro-analysis are not adequately captured by techniques designed for micro analysis. Micro-analysis of the user behaviour in our samples, focussing on a single channel at a time, thus seemed to be an approach where there would be a danger of missing some important aspects of behaviour or of failing to visualise the bigger picture of a participants' behaviour.

Proponents of schemes such as FACS do recognise the potential problems we have highlighted with these approaches. For example the Instructors Guide to FACS (Ekman, Friesen et al, 2002) recommends that coded action unit (AU) segments should be considered in context, especially of speech, head and eye movements, although only very limited means of coding of these is suggested at the current time and work is under way to evaluate when this is required. Practitioners are also directed to identify start and end points of larger 'events' that form a 'potentially meaningful unit of facial action'. Guidance is given about frequency of co-occurrence and of recognised links to underlying emotions, but ultimate decisions about the components which comprise an event are essentially qualitative.

Interesting and successful strategies for sequential analysis of interaction have been developed by Roger Bakeman and John Gottman (1995, 1997). Rather than observing models and fitting observation schemes to those models, their work involves observing interactions, defining mutually exhaustive and exclusive factors, and devising coding protocols to capture those factors and then building models to fit what is observed. For example the Specific Affect Coding System (SPAFF) was developed to assess marital interaction, which uses 5 positive codes such as interest or humour and 10 negative codes, such as disgust or anger.

The foundation of this method is the development of a catalogue of behaviour codes relevant to the particular project. Bakeman and Gottman describe these as 'the measuring instruments of observational research; they specify which behaviour is to be selected from the passing stream and recorded for subsequent study' (1997, p4). They found that having a coding strategy using of levels of coding gave more 'hooks for observers to grab' leading to collection of more reliable data.

We decided to model our approach according to the way that spoken language is analysed, using different levels of analysis. One of the best established coding systems for analyzing spoken language is the International Phonetic Alphabet (IPA). This uses sounds recognized as distinctive by native speakers of a language as units of meaning. For example, English has 20 vowel sounds and 24 consonant sounds, making 44 phonemes in all, each represented by a unique symbol. Broad phonemic transcription, is perfectly adequate for many situations and is the commonest and quickest type of transcription, used worldwide. The exact pronunciation of any phoneme will vary depending on individual variants, such as its position in a word and the influence of nearby sounds. (Compare 'l' sounds at the beginning and end of the word 'little') Variations of phonemes are called allophones. More detailed levels of coding use additional and amended symbols to show more detail about sounds. This phonological or phonetic transcription is more detailed, narrow and lengthy.

We hypothesised that similar strategies could be used to analyze the behaviour of interest to us, including non verbal behaviour and emotional expressions during HCI.

As terms within the field of 'affect' are so difficult to define, it is helpful to find uncontroversial terminology which allows us to conceptualise what we are looking for, and to build a coding catalogue to facilitate our interaction analysis. As our inductive method of identifying distinctive behaviours is analogous to the IPA, we decided to borrow from their terminology, and to identify distinctive affect related behavioural events, as ‘affectemes’. Different types of affectemes can be allocated descriptive names. As there may be variations of exact representations of these affectemes, we could then consider the variations or ‘allaffects’ displayed by our participants.

Our study aimed to identify the affective behaviour of our participants, with potential to examine events at a deep level. As our hypothesis refers to ‘more’ behaviour, we need to quantify both the number of affective events (affectemes) and their valence and intensity. We also want to assess the behavioural components of events to distinguish different types or variations of affectemes – descriped as ‘allafects’.

This initial screening approach, allows us to review our data quickly, and then spend time on in-depth analysis in the most interesting and useful parts of the data.

4.3 An Illustrative Example

When we reviewed the video footage obtained from our study it was clear that participants’ behaviour was rich in communicative, affective behaviours that were felt to indicate episodes of disaffection, tension or enjoyment. It was felt that some sort of message was being given at all times. Although events did not always contain directly affective behaviour, such as prototypical facial expression of basic emotions, all behaviour could be identified as related to affect, induced or influenced by the interaction experience.

Two one minute samples were taken from each interaction and imported into Woods and Fassnacht’s (2004) Transana video analysis tool as two set of 60 episodes (120 in all). The first sample set commenced after one minute’s interaction, to allow the participant time to ‘settle in’ to the activity, and after any influence from interaction with the research team had finished. The second sample set were of the seventh minute - towards the end of the interaction period. As context was considered critical to assessing affect, videos were edited so that the user and their screen activity could be viewed synchronously.

Samples were first generally reviewed and then meaningful episodes of affective behaviour (affectemes) were flagged and discussed. Behaviours of interest were categorised and coded. Behaviours were given intuitive names such as ‘chin dump’ to facilitate recognition and coding. Categories and codes were then developed on three levels:

- First level coding required identification of instances of communicative affective behaviours, affectemes, which could be scored for valence and intensity.
- Second level coding required analysis of incidences of whole body movements, vocalics, upper and lower facial actions, head movements, eye activity, grooming behaviours and data input activities.
- Third level analysis required identification of incidences within these second level ‘families’ for example grooming activities included biting fingers, scratching or adjusting glasses or hair.

The first level consisted of discrete episodes, but coding of instances within levels two and three could co-occur.

A coding protocol was designed to log the number of ‘affectemes’. First each video had a series of transcripts attached. Using the Transana application, transcripts can be linked by time logs to the appropriate points in the video. Clips from the video can be selected and can have keywords attached to them. Keywords families were set up ready for attachment to each transcript. The transcripts and keyword families were set up for the three levels of operation.

First Level Coding
The goal for this initial stage of interaction analysis was to identify mutually exclusive and exhaustive emotional expressions that could be counted and rated for valence and intensity. Each of the videos was viewed once to establish trends and behaviours of interest.

For the first level of analysis a transcript was attached to the video which simply marked incidences that were judged to be episodes of emotional expression. Coders were requested to time stamp any points where they perceived any communicative signals. Observers were able to identify affetemes with ease and mark their length with distinct break points. In practice coders often intuitively made inferences about participants' underlying emotional state, such as 'he looks really fed up now' and / or commented on their reasons for selecting a clip as distinctive. eg 'she did something with her eyes there'. These comments were collected along with the time codes. Each point was then revisited to establish start and end points of the affecteme and each affecteme stored as a clip. Each of our video clips was rated for valence and arousal levels using nine point scales, borrowed from Bradley and Lang's (1994) Self Assessment Manikin. It was important for us to quantify the valence and arousal levels of affectemes in order to answer our research questions.

First Level Coding - Sample Results
Initial screening identified a wide range of behaviours by individuals (see figure 2.). Occasional smiles and frowns were obvious indicators of affect, which along with occasional grimaces appeared as occasional floods of emotion crossing the face. Episodes of shifting posture – either in discomfort or in readiness for an item were also noticeable. A striking number of participants spent significant time resting their chins in their hands while they thought how to proceed. There was also noticeable 'grooming' behaviour by some participants, such as tucking their hair behind their ears or adjusting their glasses. Some participants were quite vocal, although there was no speech interface, and talked aloud to themselves, whispered clues aloud and many made non speech noises or affect bursts such as sighs, 'tutting' noises, whistling etc. Episodes of rapid blinking were also noted. Silent mouthing of clues was common, and mouth tension was displayed by fidgeting movements such as jaw grinding, lip pursing, lip licking, lip compression or mouth rinsing actions. Some common affecteme patterns were recognised and intuitively named. For example, the 'anxious peer' where participants peered at something on-screen for several moments, or the 'getting involved postural twirl' as participants shifted around and then returned to a preferred position.

……………insert Figure 2. 'Individual range of emotional expressions'…..about here

It was noticeable that judgements of affect altered over time and depending on context. For example when participants sat still without any emotional expressions while reading a new clue – this seemed comfortable and positive for a short time. However if they could not then guess the clue and proceed with the game, the same expression started to look more negative. Gaze was also significant for judgements of engagement and affect, and it was noticed that some individuals seemed to blink excessively at times during interaction.

From the second set of interaction samples (the seventh minute) we identified 401 affectemes from 60 different participants (mean 6.68). Using the Self Assessment Manikin scale, we found the mean length of an affecteme was 9.02 seconds with a mode of 3.7 seconds and a range of 0.1 to 48.3 seconds. The mean intensity rating was 4.89 with a mode of 5 and a range of 2 to 8, where 1 = high arousal and 9 = low arousal. The mean valence rating was 5.74 with a mode of 6 and a range from 3 – 8, where 1 = positive and 9 = negative.

Second level coding
For the second level of analysis, as shown in table 2, videos were assessed and behaviours recorded as a series of transcripts, using a series of seven passes looking for incidences of audible activity, whole body movements, head movements, upper face movements, lower face movements, gaze and blink patterns and activities such as keyboard entry and screen activity which were logged by spyware and pasted into the relevant transcript. Keywords were attached to instances. As our video camera was situated next to the screen, the main area of focus was the face and upper body.

---insert Table 2: Level 2 Transcripts and keywords used in this analysis, about here---

By reviewing each one minute video sample a number of times, and recording a series of different aspects of behaviour as separate transcripts, it was possible to look for one behaviour at a time, in a systematic and thorough way. Observers were able to identify different types of affectemes. Sometimes one affecteme contained several different modes, for example displaying a chin dump, a peer, an episode of rapid blinking and a sigh all within one short episode.

Second Level Coding - Sample Results
Content varied, for example during the first set of interaction samples (taken after one minute) we noted:

- 53 episodes of 'chin dumps' where participants rested their head in their hands, (mean 0.88 per participant). The total duration was 581 seconds (mean duration 9.68 seconds per participant);
- 287 episodes of postural shifts (mean 4.78 per participant.) The total duration of these was 517 seconds (mean 8.62 seconds per participant);
- 29 episodes of grooming in the form of adjusting hair, clothing or spectacles etc. (mean 0.48 per participant) with total duration of 65 seconds (mean duration 1.08 seconds per participant);
- 69 vocalic instances, such as brief sighs groans or laughs (mean 1.15 per participant);
- 803 instances of blinking, (mean 13.38 per participant).

There were individual differences between participants, so that not all displayed all affecteme components:

- All 60 participants used postural shifts (duration range 1-41 seconds, mean 8.62);
- 32 participants used chin dumps (duration range 1-47 seconds, mean 18.16);
- 31 participants used vocal sounds, mainly affect bursts (range 1-6 instances, mean 2.23);
- 12 participants used grooming behaviours (range 1-5 instances, mean 2.42).

Third Level Coding
For further analysis at the third level of coding, when required, our top-down approach allowed us to code variations on an affecteme and episodes could be further categorised within the keyword family, so that the type of smile, or exact nature of a chin dump could be transcribed by adding appropriate keywords from that family. For example a chin dump might involve one or both hands; contact with the cheek or the chin, contact of whole hand palm or fingertips; covering of mouth, mouth and nose or neither. For the purposes of our analysis these could all be classed as the same behaviour, although fine details might vary, as shown in Figure 3.

….insert Figure 3. 'Examples of allafects of a chin dump' about here…………………

Reliability of the Method
Our research question requires us to identify emotion related episodes of behaviour (affectmes) and rate the valence and arousal levels displayed by our participants during them. We need to check that different raters identify episodes and ratings consistently. We followed three phases to establish agreement. First we examined a small sample of data in some depth to establish our initial criteria and identify affectemes. Secondly we ran a trial study to code data for valence and arousal, comparing results of pairs of coders, to inform what additional training coders would require and what refining our criteria and procedures needed, as calculations for each coder pair can be a valuable diagnostic tool during coder training (Lombard, Snyder-Duch et al, 2005). Thirdly we ran a pilot coding exercise to establish reliability was satisfactory across multiple coders.

Reliability Agreement - First stage
For the first stage a one minute pilot sample was rated by three coders (including the researcher) and differences were noted and discussed. It proved intuitive to identify segments where a discrete message or episode of affect was perceived, although sometimes identification of onset and offset times varied by a few seconds, as identified by different coders. In some instances one coder would initially identify fewer segments as occurring than another, although when segments were discussed, then there was agreement that two segments could be construed as parts of a larger meaningful segment. We achieved percentage agreement levels of over 90%, and Krippendorff's alpha .9740 for agreement on identifying affective segments. Episodes of affective behaviour were discussed and named. Using Transana qualitative analysis software video clip collections were built, which allowed clips to be viewed in sequence, for example all 'chin dumps' can be reviewed, to check that they are all instances of the same behaviours.

Once segments were identified, than each segment was then rated for valence and intensity. These ratings were transcribed for each second of activity. Confusion matrices were quickly drawn to provide quick visual comparisons of results. This showed that although ratings were not exactly the same, ratings were usually in the same quadrant.
The Self Assessment Manikin proved very helpful and intuitive to use as an assessment tool. Using this 9 point scale, ratings were often similar, falling within the same range, but not always identical, as shown in Table 3.

………………………..insert Table3. 'Example of data' about here…………………

In order to assess the reliability of our coding of content analysis, we need to assess more than correlation or analysis of variance values. We need to assess the intercoder agreement, that is the extent to which the different judges tend to assign exactly the same rating to each object. Lombard, Snyder-Duch et al (2002) reviewed literature regarding intercoder reliability and found poor use and reporting of methods. They developed useful guidelines for appropriate procedures, which we have followed, which recommend reporting actual intercoder agreement for variables on coded values which should be the basis for assessment used (even if, as with our data, similar rather than identical values 'count').

We selected several indices to measure agreement. First we considered agreement between pairs of raters using percentage agreements, and Cohen's Kappa to account for agreement expected by chance between pairs of raters. Despites some drawbacks, this is regarded as the "measure of choice" for rating behaviour by Bakeman (2000). We then used Krippendorff's alpha to consider coefficients between multiple raters.

Reliability Agreement - Second stage

For the second stage, a representative sample of our entire data was selected to be tested for pilot reliability testing consisting of 6 one minute samples of data (10% of total data). These samples are a subset of our main data, selected to represent the main conditions of our experiment and to represent both male and female participants. We used 5 coders including the researcher to code these samples. Each coder received a 10 minute training session, reviewing a one minute sample with the researcher and discussing the coding criteria. At this stage we compared results between pairs of coders for their ratings of each 60 second sample, giving six comparisons per pair of coders per variable.

We hoped to achieve percentage agreements of 75% or above as this is a liberal agreement index (Lombard, Snyder-Duch et al 2005). Using the more stringent and conservative Cohen's Kappa or Krippendorff's alpha we would be satisfied with relatively low coefficients of .70 or greater as this is an exploratory study (Lombard, Snyder-Duch et al 2005).

We used Bakeman's (1995) Generalised Sequential Querier, 'GSEQ' software to calculate percentage agreements and Cohen's Kappa between pairs of coders and we used the Statisitcal package for social sciences software, (SPSS) with a macro for Krippendorff's alpha (Hayes, 2005) with which computes alpha for any level of measurement and any number of judges.

Lombard, Snyder Duch et al (2005) state that "there is general consensus in the literature that indices that do not account for agreement that could be expected to occur by chance (such as percent agreement…..) are too liberal (i.e., overestimate true agreement), while those that do can be too conservative." Cohen's kappa and Krippendorff's alpha do not allow any credit for ratings which are close, but not exactly the same, although in this study, if one rater put valence at 7 and one at 8 that might be considered close enough. Although our percentage agreements were good, when the more stringent tests were used, the reliability scores were lower.

For valence ratings, our percentage agreements ranged from 83.33 to 100% with a mean of 92.2%. Kappa agreements ranged from 0.11 to 1.0 with a mean of 0.58. Krippendorff's alpha was 0.2301.

For arousal ratings our percentage agreements ranged from 83.33 to 97.78% with a mean of 89.8%. Kappa agreements ranged from 0.13 to 0.86 with a mean of 0.45. Krippendorff's alpha was 0.0232.

As these results did not meet our agreement criteria, we examined the results for pairs of coders in detail. It was interesting to note that there was higher agreement for valence than for arousal ratings, and that there was higher agreement between same sex pairs of raters, particularly when female and particularly when rating the opposite sex.

Reliability Agreement - Third Stage
In view of the stage two results we decided to test for reliability using four female raters and a longer training session lasting one hour, during which six one minute samples (3 male and 3 female) were examined in detail. Each rated the 6 samples used previously, following the same procedure as before. Ratings were closer than in the previous pilot. Whereas ratings had sometimes differed by up to 4 points on the rating scale in the pilot ratings, they were now within 2 points in all cases.

We used 360 seconds of data from the 6 one minute samples for valence, to compute Krippendorff's alpha and repeated the exercise using arousal data, so that there was one comparison for each set of coders for each variable. Krippendorff's alpha met our criteria for agreement with valence rating agreement of 0.7836 and arousal rating agreement of 0,7165.

5. Conclusion and Discussion
This paper has presented a novel method for coding user emotional expression during interactive experiences. The method was developed as a result of our work to analyse the data from an experiment on human-computer interaction with future affective systems. During this analysis we recognised shortcomings of existing coding approaches and hypothesised that a technique derived from those used in interaction analysis and spoken language transcription might be beneficial in this context. The method involves identifying affectemes (which are analogous to phonemes in language transcription). Finer grained analysis can also be performed to identify variations within any one affecteme (allafects which are analogous to allophones in language transcription).

This method enabled us to identify and code the behaviours we were interested in, in context, during HCI and to consider their components. Although observational schemes have limitations, this method has provided a structured way to assess observational data and to objectify and quantify it.

Experience with using this coding scheme suggests that it is straightforward to apply, is quicker than alternatives such as FACS and is easy to learn. We also found that it provided useful insights into the patterns of emotional expressions displayed at the user interface. Preliminary evidence on the reliability and validity of the method is also encouraging. Further work is needed to explore these issues in more detail.

Our method succeeded in showing differences between affectemes. By adding further levels of coding more sensitivity could be achieved. For example it would be useful to use this method to identify segments for detailed coding using established coding schemes, such as FACS.

Future analysis needs to consider sequential patterns in the presentation of affecteme components and the onset, peak and offset of actions, and the likely transitions between levels of intensity and valence.

It is tempting to ascribe underlying emotion to affectemes. However the same emotional expression may have different underlying meaning. It might be useful to adopt a structured framework to try to appraise meaning, in context, and this is another direction for future work.

The main contribution of this paper is to provide HCI researchers with an approach to coding affective and expressive behaviours during interactions with computers. While the basis for the technique is not new, drawing as it does on well established techniques from interaction analysis, this type of approach may be unfamiliar to some HCI practitioners and its application in this domain is novel. In addition we have provided a new vocabulary for discussing units of affective or communicative behaviour, drawing upon the well known distinctions made in spoken language transcription. We hope that providing this vocabulary will facilitate further research in this area by providing researchers with a common language with which to describe their findings.

We have also begun to identify some of the affectemes which can be observed during human-computer interaction, (and which may, or may not be specific to our puzzle solving task), such as the ‘fed up chin dump’, the ‘anxious peer’ and the ‘getting involved postural twirl’. Work on identifying the affectemes of HCI is in its very early stages but it may eventually be possible to define a dictionary of meaningful units of affective behaviour, or HCI affectemes, relevant to different domains. By describing our work here and presenting a vocabulary for such analysis, we hope to encourage other researchers to contribute to building up such knowledge.

5.1 Comparison to FACS

FACS is probably the best known and most well researched coding scheme for affect. We recognise that FACS is an extremely sophisticated tool, backed by extensive research and providing highly detailed data on human facial movement. Such data can be extremely valuable in many areas of study. However, we do argue that FACS may not always be the most appropriate tool to use to study human expressions during human-computer interaction. The focus within FACS is on the analysis of very fine levels of detail about the movements of facial muscles. In many practical applications this level of detail may simply not be needed and may not therefore justify the time cost of the coding. Ekman, Friesen and Hager recognise this themselves, stating that:
‘FACS is a time consuming procedure for measuring facial movement. Many of the distinctions between AUs or the full description and timing of complex combinations may be irrelevant to certain questions. It is hoped that future empirical data will identify distinctions and procedures which can be collapsed.’ (Ekman, Friesen et al, 2002, p181).

Our argument is based around a number of different factors, summarised in table 4:
(a) time to learn and apply the technique
(b) the focus of the analysis (e.g. mirco vs. macro level analysis)
(c) theoretical commitment associated with the method
(d) constraints on images to be coded
(e) dealing with inter-personal and intra-personal differences

We hope to demonstrate that the Affecteme approach provides tangible benefits in each of these areas. We also argue that where FACS is used, the Affecteme approach can be a useful first step in narrowing down the samples to be analysed in further depth using FACS coding. Different techniques should therefore be seen as complementary and not as competing.

………insert table 4. 'Comparison of FACS and Affectemes' about here

5.2 Complications of multimodal expressions

The unpredictability of humans and the multimodal nature of communication creates difficulties for any automatic detection system that sets out to interpret emotional expressions without situational awareness. For example a number of our participants obscured some of their facial actions with their hairstyles or by wearing hats or glasses, chewing gum or peering partly off camera at times.

When the same expression on a face is seen from different angles it can appear to be smiling or unhappy. Mixed messages between modes occur and sometimes are used as a communication strategy, for example irony is often conveyed by a mismatch between tone of voice and choice of words. Some facial actions can be used for different purposes. For example there is a typical expression of grief – known as 'Darwin's grief muscle' as a result of Darwin's book (Darwin 1872) on the expression of emotion. Bakeman and Gottman (1997) noted that this is the exact expression typically used by the comic actor Woody Allan to signal delivery of a joke. It is not possible to assess the emotional expression unless you make an appraisal of the intent behind the emotional expression.

Duplication of meaning in multiple modes occurs and is know as 'redundancy'– but is a useful strategy in situations where some modes may be lost or poorly perceived, such as using gestures to assist understanding in a noisy environment, or tone of voice to understand valence when face is not visible. Modes are not only used to communicate affect, but can add emphasis or extra information, for example Baveles (1996) describes a participant's use of gesture in a narrative about a dynamic event (a 'near-miss' car crash), where spoken information is enhanced by gestures to symbolize movements and add information about direction, position and force and the consequent intensity of the emotional experience.

Emotional expressions may also be used to convey an attitude or judgement rather than an underlying feeling, for example a frown to show disapproval rather than real anger. Emotional expressions can be at odds with underlying emotional experience, for example a smile faked, for personal or social reasons. Inhibition can speak, and in a certain context (e.g. where animation is expected) a blank face gives a powerful message. Emotional expressions may be under conscious or unconscious control and may be 'leaked' unwittingly. Theories such as that described by Kaiser, Wherle et al, (1998) link emotion-antecedent cognitive appraisal and associated action tendencies, the expression of which we may inhibit or regulate. Baveles (1996) suggests that many widely reported beliefs about the meanings of non-verbal communication are based on very flimsy or non-existent evidence.

It may be that certain modes are highly significant indicators for certain emotions, but if so we have not yet reliably mapped them. In the absence of sure indicators, we must consider information from all modes. In face to face situations, we constantly make judgements about people's affect by using all of our senses, and our situational knowledge of place, time and shared experiences.

5.3 New multimodal rules

With new technologies, we adapt our communication skills. With text and email, new conventions are developing, for example Wallace (1999) describes the use of 'emoticons' to add information about emotional states to email messages. There are relatively few studies reported that describe user's 'normal' non verbal behaviour during human-computer interaction (HCI). In an interesting series of experiments, Reeves and Nass (1998) have found we tend to treat new media like real people and real things. However there may be costs in learning new communication strategies. Keates and Robinson (1999) found systems designed for gestural input held promise, but imposed extra cognitive and physical loads on users. If such developments are to succeed we need to recognise and harness behaviours that users exhibit naturally, and without effort.

5.4 Mapping to underlying emotions

Affectemes make no claim to map to underlying emotional experiences, although they may do so. Emotional expressions have a complex relationship to underlying emotional experiences. Ekman (2003) comments on individuals who take pleasure in fear, for example in the context of a scary movie. An example within our dataset, was when one participant smiled, seemingly not to indicate positive affect as it followed a series of bad guesses and errors in entering an answer to the word puzzle. It seemed more likely to be a 'wry' smile, as the participant laughed at himself and his own mistakes, possibly to diffuse any tension generated by the failures. As another example participants used 'tutting' sounds sometimes directed at themselves when they made and error and sometimes directed at the game when it presented them with an unwanted response. It is essential to consider the interactive discourse between participant and application to have any understanding of meaning. Raters intuitively used emotion related words to describe behaviours that were sometimes linked to perceived emotional states (for example 'he looks happy') or sometimes linked to appraisal of the interaction, (for example 'he wants another clue'). A possible extension of the work would be to consider a structured appraisal analysis framework to link affectemes to underlying meaning and intent, whether emotional or not.

5.5 Reliability of coding

Establishing reliability of coding is complex and there is no ideal method recommended that gives credit for close rating scores, when using Likert type scales. Several refinements of procedures were needed in order to bring agreement up to an acceptable level. Whilst we found the Self Assessment Manikin was a very useful tool, improved reliability may be achieved by devising a less sensitive scale, collapsed to fewer points. We found differences between pairs of raters dependent upon gender of the raters and participant samples. In the early stages of agreement rating, three of our raters were female and two male and the subject sample were 3 male and 3 females. There was higher agreement between pairs of raters of the same gender, particularly when they were rating the opposite gender. Female agreement was greater than agreement between males. This is consistent with the literature on gender differences in the perception of affect. For example from an early age and across many cultures, females have greater ability than males to perceive facial expressions of emotion (Elfenbein Marsh et al, 2002). Males have particular inability to recognise females' negative displays and ability to recognise facial expression may also be affected by menstrual and testosterone cycles (Goos & Silverman, 2002). This might have important implications for standardised collection of observational data on affect, and larger studies are needed to investigate this.

5.6 Retrospective comparison

We carried out retrospective walkthroughs to see how our participants rated their behaviour in comparison to objective raters. We found good agreement between participants and raters about onset and offset of events. Rating agreements were often just one or two points apart, with valence more easily agreed than arousal. There was closer agreement about ratings at extremes, for example very positive or negative, but less agreement about more neutral states. There was also some interesting discussion about appraisal of the participants' behaviour. Different uses of affective terminology for emotional states sharing the same general valence was striking. For example, on one event where the rater had described the participant as 'anxious', the participant described himself as 'confused' and 'angry'.

5.7 Relevance for designing affective systems

Observers tend to agree when a communicative or affective event is taking place, and naturally use affective vocabulary to describe those events. They are less able to articulate the exact reason for their observations, or to describe all aspects of the behaviour, choosing to extract certain salient signals, which may be modified or have nuance added by signals from other modes. By analysing the lower level behaviours that participants demonstrate, or components of affectemes, it may be possible to extract salient features of particular affectemes that would be useful to developers of recognition systems. From our study so far we have identified a number of affectemes, that are commonly exhibited by our range of participants, in our interactive puzzle solving game context.

We have a large amount of data and there is further scope to examine and analyse it. Onsets, offsets and peaks of affetemes could be examined in terms of components and then further sequential analysis may help identify features that could be of use in identifying probability of a particular action representing a particular affecteme. As well as being potentially useful for analysing user affective behaviour during human-computer interaction, it is also possible that the type of results obtained using the affecteme approach may prove useful to those developing algorithms to detect and recognise emotional expression at the human-computer interface. Development of such systems might be accelerated by concentrating on the recognition of affectemes, or components of them, relevant to particular applications, rather then trying to capture every detail of human emotional behaviour.

6. Acknowledgements

This research is supported by the EPSRC, grant reference R81374/01.

7. References

ARGYLE, M., 1988, *Bodily Communication*, 2nd Edition. New York: Methuen.

AXELROD, L., HONE, K., 2005, E-motional Advantage: Performance and Satisfaction Gains with Affective Computing. *Proceedings of CHI 2005 'Technology, Community, Safety' Conference on Human Factors in Computer Systems*, Portland, USA

BAKEMAN, R., 2000. Behavioral observation and coding. In H. T. Reis & C. M. Judge (Eds.), *Handbook of research methods in social and personality psychology* (pp. 138-159). New York: Cambridge University Press.

BAKEMAN, R., GOTTMAN, J. M., 1997, *Observing Interaction: An introduction to sequential analysis*. Cambridge University Press

BAKEMAN, R, QUERA, V., 1995, *Analyzing Interaction: Sequential analysis with SIDS and GSEQ*. Cambridge University Press.

BAVELAS, J. B., 1996, *Debunking Body Language: new research on non-verbal communications.* Lecture given at University of Victoria available at http://novaonline.nv.cc.va.us/eli/spd110td/interper/message/linksnonverbal.html

BLYTHE, M., MONK, A.F., OVERBEEKE, K., WRIGHT, P.C., (Eds.) 2003, *Funology: From Usability to Enjoyment* (Human Computer Interaction Series, 3) Kluwer Academic Publishers.

BRADLEY, M. M., LANG, P. J., 1994, Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *J Behav Ther Exp Psychiatry* 25(1), 49-59.

BUNTING, M., 2004, *Willing Slaves: How The Overwork Culture Is Ruling Our Lives* HarperCollins.

CORNELIUS, R. R., 1996, *The Science of Emotion*. Prentice-Hall Inc.

CUTLER S. E., LARSEN R. J., BUNCE S. C.**,** 1996, Repressive coping style and the experience and recall of emotion: a naturalistic study of daily affect. *J Pers*. Jun;64 (2): 379-405.

DARWIN, C., 1872, *The Expression of the Emotions in Man and Animals*. New York: D. Appleton & Company. 176-195

DUTTON, D., ARON, A., 1974, Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, 30, 510-517.

EKMAN, P., 2003, *Emotions Revealed: Recognising faces and feelings to improve communication and emotional life.* Times Books.

EKMAN, P., FRIESEN, W. V., HAGER, J. C., 2002 *Facial Action Coding System: Investigator's Guide.* Research Nexus

EKMAN, P., ROSENBERG, E., (Eds.) 1998, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)* (Series in Affective Science) Oxford University Press Inc, USA.

ELFENBEIN, H.,MARSH, A., AMBADY, N., 2002 Emotional Intelligence and the Recognition of Emotion from Facial Expressions. in *The Wisdom in Feeling:Psychological Processes in Emotional Intelligence* Lisa Feldman Barrett & Peter Salovey, Eds Guilford Press

FANG, X, SALVENDY, G., 2001, A personal perspective on behaviour and information technology: a 20-year progress and future trend. *Behaviour and Information Technology*, 20(5), 357-366.

GIMSON, A. C., CRUTTENDEN, A., (Ed.) 2001, *Gimson's Pronunciation of English.* Hodder Arnold.

GOOS, L., SILVERMAN, I., 2002 Sex Related Factors in the Perception of Threatening Facial Expressions *Journal of nonverbal Behavior*, 26 (1): 27-41, Spring 2002

HARWOOD, N. K., HALL, L. J., SHINKFIELD, A. J., 1999, Recognition of Facial Emotional Expressions From Moving and Static Displays by Individuals with Mental Retardation. *American Journal on Mental Retardation.* Vol. 104 (3), 270–278.

HAYES, A., 2005 Website – SPSS macros http://www.comm.ohio-state.edu/ahayes/

HOCHSCHILD, A. R., 1983 - *The Managed Heart: Commercialization of Human Feeling.*

HUTCHINSON, G., HUTCHINSON, A., ANDERSON, D., 1987, Labanotation: The System of Analyzing and Recording Movement. Routledge, Inc.

IPA website available at http://www.arts.gla.ac.uk/IPA/index.html

ISLE, 2002, *Survey of Multimodal Coding Schemes and Best Practice.* Natural Interactivity and Multimodality Working Group Deliverable D9.1. Available at http://isle.nis.sdu.dk/

KAISER, S., WEHRLE, T., SCHMIDT, S., 1998, Emotional episodes, facial expressions, and reported feelings in human-computer interactions. In A. H. Fischer (Ed.), Proceedings of the Xth Conference of the International Society for Research on Emotions(pp. 82-86).Würzburg: ISRE Publications.

KAISER, S., WHERLE, T., 1994, Emotion research and AI: some theoretical and technical is sues. *Geneva Studies in Emotion and Communication.*

KAISER, S. & WEHRLE, T., 1992. Automated coding of facial behavior in human-computer interactions with FACS, *Journal of Nonverbal Behavior, 16* (2), 67-83.

KEATES, S., ROBINSON, P., 1999, Gestures and multimodal input. *Behaviour & Information Technology*, 18(1) 36-44.

KIPP, M., 2004, *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*, Boca Raton, Florida: Dissertation.com. http://www.dfki.de/~kipp/anvil/

KIRSCH, D., 1997**,** *The Sentic Mouse: Developing a tool for Measuring Emotional Valence,* MIT Technical Report TR-495 available at http://vismod.media.mit.edu/tech-reports/TR-495/

LOMBARD, M., SNYDER-DUCH, J., BRACKEN, C., 2005 Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects http://www.temple.edu/mmc/reliability/

LOMBARD, M., SNYDER-DUCH, J., BRACKEN, C., 2002 Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability *Hum Commun Res* 2002 28: 587-604

MADDOCK, R.J., GARRETT, A. S., BUONOCORE, M. H., 2003, Posterior Cingulate Cortex Activation by Emotional Words: fMRI Evidence From a Valence Decision Task. *Human Brain Mapping* 18:30–41

MAYNE, T. J., BONANNO, G. A., 2001, *Emotions: Current Issues and Future Directions.* Guilford Press.

METTE, T., POSAMENTIER, M. T., ABDI1, H., 2003, Processing Faces and Facial Expressions. *Neuropsychology Review*, Vol. 13(3).

MINSKY, M., 2004, *The Emotion Machine* (draft) available at www.media.mit.edu/~minsky/

NORMAN, D. A., 2004, *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books

PICARD, R.W., 1997, *Affective Computing*. Cambridge,MA: The MIT Press.

REEVES, B., NASS, C., 1998, *The Media Equation: How people treat computers, television and new media like real people and places.* CSLI Publications.

SCHACHTER, S., SINGER, J.E., 1962, Cognitive, social, and physiological determinants of emotional states. *Psychological Review 69*, 379-399.

SCHEIRER, J., FERNADEZ, R., KLEIN, J. AND PICARD, R. W. 2002, Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, 14(2), 93-118.

SCOLLON**,** C. N., SHIGEHIRO, O., BISWAS-DIENER, R., 2004, Emotions across cultures and methods *Journal of Cross-Cultural Psychology*; 35: 304-326.

ToBi website available at http://www.ling.ohio-state.edu/~tobi/

WALLACE, P., 1999, The Psychology of the Internet. Cambridge University Press.

WARD, R.D. AND MARSDEN, P.H., 2003, Physiological responses to different web-page designs. *International Journal of Human-Computer Studies*, 59(1/2), 199-212.

WARD, R.D. and MARSEN, P.H. (in press) Affective computing: problems, reactions and intentions. To appear in Interacting With Computers.

WOODS, D. K., FASSNACHT, C, 2004 Wisconsin Centre for Education Research, TRANSANA website http://www2.wcer.wisc.edu/Transana

Table 1 Participant groups – 2X2 factorial design

	Application appeared to react to emotional expressions	Standard application
Participants told application was affective	Group 1	Group 2
Participants not told application was affective	Group 3	Group 4

Table 2: Level 2 Transcripts and keywords used in this analysis

Transcript	**Keywords-affectemes**	**Keyword families – examples of allaffects**
1 audible activity	speech	
	whisper	
	affect burst	groan, whistle, tut, sigh, indrawn breath, snort.
	extraneous noises	Foot tap, finger tap, door bang
	data entry	keyboard, mouse click
2 whole body movement	shifts	large shift, small shifts, lean back, postural twirl
	tension	hunched shoulders
	grooming	tuck hair, shunt glasses, bite finger/thumb, scratch
	chin dump	L hand, R hand, both hands, mouth covered, fisted, fingers spread
3 head movements	gesture	nod, shake
	peer	distant, close
	aspect shifts	tilt, turn, L, R, up, down
	chin move	tuck, thrust
4 upper face	brow-raise	bilateral, unilateral L, unilateral R
	frown	slight, deep
	nose	wrinkle, flare
5 lower face	smile	PanAm, zygomatic
	mouthing words	
	fidgets	compress, pursing, rinsing, jaw grind
6 gaze and blinks	eye shifts	flashbulb eyes, narrowed, closed
	screen attention shifts	on screen, off screen, scanning, L, R, up, down
	blinks	
7 keyboard, mouse and on-screen activity	data entry	Keystrokes, mouse clicks
	on-screen activity	Picture clue, short text clue

Table 3. Example of data

Second Number	**91**	**92**	**93**	**94**	**95**	**96**	**97**	**98**	**99**	**100**
Rating given by Rater 1	7	7	7	7	7	7	4	4	4	4
Rating given by Rater 2	7	7	7	7	7	7	4	4	4	4
Rating given by Rater 3	7	7	7	7	6	6	6	6	6	6
Rating given by Rater 4	6	7	7	7	5	5	5	5	5	5

Table 4. Comparison of FACS and Affectemes

Variable	FACS	Affectemes
Approach	Micro analysis – bottom-up Sophisticated tool	Macro analysis – top-down 'Quick and dirty' method
Time to learn	100 hours +	1 hours
Time to apply	1 hour per minute of video +	5-10 minutes per minute of video, depending on level of analysis required.
Automation	Prototype already developed to recognise facial actions.	Not automated and would be difficult to do so as complex and multimodal expressions involved
Expertise	Desirable for raters to become highly skilled 'facial athletes', take an exam and practise regularly to maintain reliability	Inductive and intuitive for raters. Need to plan together and calibrate for a particular project.
Analysis tools		Data can be organised with shareware qualitative analysis applications such as Transana (Woods + Faasnacht, 2004) or Anvil (Kipp, 2004).
Detail	Very fine level	Different levels depending upon relevance
Consistency	Very consistent	Can be adapted to look for different behaviours depending on relevance
Flexibility	Not flexible	Flexible
Multimodality	Not multimodal – have to relate to context during final stages of analysis advise that 'other sources of information should be considered'.	Multimodal aspects incorporated from the start
Fit to context	Has to be related to context during final stages of analysis	Context considered from the start
Map to underlying emotions	Good evidence for mapping of some combinations of action units	No attempt to map to underlying emotions. Possibility of adding appraisal framework to assess communicative intent.
Theoretical basis	Basic emotions	No strong link to one theory – but can adapt so could link to identify expressions of basic emotions and or to dimensional model also.
Events	Delineation at micro level can be hard to identify	Delineation intuitive
Resulting data	Essentially Quantitative	Essentially Qualitative
Constraints	Full clear view of face essential	Capitalises on redundancy of multimodal expressions so human foibles such as leaning off camera, wearing baseball hats and chewing gum do not necessarily rule out possibility of coding.
Inter and intra personal differences	No consideration of differences in or between subjects during coding.	Rater can consider differences, such as habitual facial expression, during coding

Figure 1. Word-ladder puzzle task.

Figure 2: Individual range of emotional expressions

Figure 3: Examples of allafects of a 'chin dump'

Appendix E. Advert, Game clues, Participant Instructions and Consents, Pre and Post Experiment Questions.

ADVERT

COMPUTER USABILITY EXPERIMENT running from now until 31st May

RING NOW

Make an appointment to come along to the Department of Information Systems and Computing (Pink portacabin building next to St John's building, Uxbridge campus) and take part in an experiment to see how usable you find a short computer puzzle. It will take about 1 hour to play the game and answer some questions afterwards.

If you would like to take part ring Lesley on xxx xxxx xxxx or email l.axelrod@brunel.ac.uk

Computer usability experiment
Ring Lesley on 077 xxxx xxxx
Email v

Computer usability experiment
Ring Lesley on 077 xxxx xxxx
Email l.axelrod@brunel.ac.uk

Computer usability experiment
Ring Lesley on 077 xxxx xxxx
Email l.axelrod@brunel.ac.uk

Computer usability experiment
Ring Lesley on 077 xxxx xxxx
Email l.axelrod@brunel.ac.uk

Computer usability experiment
Ring Lesley on 077 xxxx xxxx
Email l.axelrod@brunel.ac.uk

Computer usability experiment
Ring Lesley on 077 xxxx xxxx
Email l.axelrod@brunel.ac.uk

Computer usability experiment
Ring Lesley on 077 xxxx xxxx
Email l.axelrod@brunel.ac.uk

Computer usability experiment
Ring Lesley on 077 xxxx xxxx
Email l.axelrod@brunel.ac.uk

Brunel University,
Department of Information Systems and Computing
St John's Building, Uxbridge Campus,
Uxbridge UB8 3PH
01895 274000

Explanation Condition 1

DISC, Computer Usability Project

This is part of a three year project on computer usability of applications that can recognise and respond to some of the emotions of the people using them. It is funded by the EPSRC. The project is directed by Kate Hone, Senior Lecturer in Brunel's Department of Information Systems and Computing. The experiment will be conducted by Lesley Axelrod. You can contact the team at the above address.

This purpose of this experiment is to explore how usable people find computer systems, that can recognise and respond to the emotions of the person using them. During the course of the experiment you may be recorded on tape or film, and you will be asked to provide information about yourself. These recordings and information will be used to help analyse how usable you find the computer system. They will be kept securely according to Brunel University code of practice. They will only be used by the project team. They will be destroyed on the completion of the project, unless you give your permission for their further use. You have a right to privacy and your name will not be revealed in any way. If at any time you want to discontinue the experiment, you are free to do so.

1. You will be asked to complete a questionnaire giving some basic information about yourself and your computer use.

2. You will be asked to attempt some simple puzzles using the computer, which may respond to some of your emotions.

3. You will be asked to complete a questionnaire giving more information about yourself and your views on the usability of the system.

This experiment has been explained to me by Lesley Axelrod.
I give permission for recordings to be made of me during the experiment and I understand the recordings will only be used by the project team, and destroyed at the end of the project, unless I give further permission for their use.

Signed……………………………………………………Date…………………………

I give further permission for clips from any recordings made of me during the experiment to be used in any future publications or presentations by the research team.

Signed……………………………………………………Date…………………………

Brunel University,
Department of Information Systems and Computing
St John's Building, Uxbridge Campus,
Uxbridge UB8 3PH
01895 274000

Explanation Condition 2

DISC Computer Usability Project

This is part of a three year project on computer usability funded by the EPSRC. The project is directed by Kate Hone, Senior Lecturer in Brunel's Department of Information Systems and Computing. The experiment will be conducted by Lesley Axelrod. You can contact the team at the above address.

This purpose of this experiment is to explore how usable people find computer systems. During the course of the experiment you may be recorded on tape or film, and you will be asked to provide information about yourself. These recordings and information will be used to help analyse how usable you find the computer. They will be kept securely according to Brunel University code of practice. They will only be used by the project team. They will be destroyed on the completion of the project, unless you give your permission for their further use. You have a right to privacy and your name will not be revealed in any way. If at any time you want to discontinue the experiment, you are free to do so.

1. You will be asked to complete a questionnaire giving some basic information about yourself and your computer use.

2. You will be asked to attempt some simple puzzles using the computer.

3. You will be asked to complete a questionnaire giving more information about yourself and your views on the usability of the system.

This experiment has been explained to me by Lesley Axelrod.

I give permission for recordings to be made of me during the experiment and I understand the recordings will only be used by the project team, and destroyed at the end of the project, unless I give further permission for their use.

Signed……………………………………………………Date…………………………

Brunel University,
Department of Information Systems and Computing
St John's Building, Uxbridge Campus,
Uxbridge UB8 3PH
01895 274000

Debrief Condition 2

DISC Computer Usability Project

The explanation given to you at the start of this experiment may not have been complete. This experiment is not only looking at the usability of the game. We are also studying how people interact with computer systems that can respond to emotions. For half the participants that take part, the word-ladder game is manipulated by the researchers, so that different clues and messages appear, so that the system appears to respond to some of your emotions. We did not want to alter people's behaviour by fully telling them in advance, what to expect. We hope you will accept out apologies if you were initially misled.

I give permission for my data to be included in the experiment.

Signed……………………………………………………

Date………………………………………………………

I give further permission for clips from any recordings made of me during the experiment to be used in any future publications or presentations by the research team.

Signed…………………………………………………….

Date………………………………………………………

Brunel University,

Department of Information Systems and Computing
St John's Building, Uxbridge Campus,
Uxbridge UB8 3PH
01895 274000

Explanation Experiment Two, Condition 2

DISC, Computer Usability Project

This is part of a three year project on computer usability of applications that can recognise and respond to some of the emotions of the people using them. It is funded by the EPSRC. The project is directed by Kate Hone, Senior Lecturer in Brunel's Department of Information Systems and Computing. The experiment will be conducted by Lesley Axelrod. You can contact the team at the above address.

This purpose of this experiment is to explore how usable people find computer systems, that can recognise and respond to the emotions of the person using them. During the course of the experiment you may be recorded on tape or film, and you will be asked to provide information about yourself. These recordings and information will be used to help analyse how usable you find the computer system. They will be kept securely according to Brunel University code of practice. They will only be used by the project team. They will be destroyed on the completion of the project, unless you give your permission for their further use. You have a right to privacy and your name will not be revealed in any way. If at any time you want to discontinue the experiment, you are free to do so.

1. You will be asked to complete a questionnaire giving some basic information about yourself and your computer use.

2. You will be asked to attempt some simple puzzles using the computer, which may respond to some of your emotions. You will be asked to use some particular emotional expressions (explained and demonstrated by the researcher) during the game.

3. You will be asked to complete a questionnaire giving more information about yourself and your views on the usability of the system.

This experiment has been explained to me by Lesley Axelrod.
I give permission for recordings to be made of me during the experiment and I understand the recordings will only be used by the project team, and destroyed at the end of the project, unless I give further permission for their use.

Signed……………………………………………………

Date………………………………………………………

I give further permission for clips from any recordings made of me during the experiment to be used in any future publications or presentations by the research team.

Signed…………………………………………………….

Date………………………………………………………

Game instructions

You are going to attempt a couple of word puzzles called 'Word ladders'.
You will be given the first and last words on the ladder.

For example

1.	HEAD
2.	
3.	
4.	BEER

You have to type in a word in the second 'rung' of the ladder which is different from the first word by one letter. You will be given a clue to help you. If you want an additional clue press the 'help' button and another clue will appear in a new window. When you have seen the clue, click on the 'close' button to remove the new clue window and continue with the word ladder.

For example

1.	HEAD
2.	HEAR
3.	
4.	BEER

Then the third 'rung'

1.	HEAD
2.	HEAR
3.	BEAR
4.	BEER

Mistakes

If you type in the wrong word it will not be accepted and you will have to try that 'rung' again.

Using the keyboard.

- You can type in your answers
- You can use the tab key to move on to the next rung.

Using the mouse

You can use the mouse:

- to reposition the cursor
- to click the 'NEXT' button for the next clue
- to click for 'HELP'.

YOU WILL BE WORKING AGAINST THE CLOCK. GOOD LUCK.

GAME CLUES

GAME 1.

GAME 1	target	Text Clue 1	Text Clue 2	Picture Clue
Starting word	four			
Step 1	foul	offensive to the senses, full of dirt or mud, morally or spiritually odious, notably unpleasant or distressing, constituting an infringement of rules in a game or sport, encrusted, clogged, or choked with a foreign substance	horrible	a referee holding up a red ticket
Step 2	fool	a person lacking in judgment or prudence, a retainer formerly kept in great households to provide casual entertainment and commonly dressed in motley with cap, bells, and bauble, a harmlessly deranged person or one lacking in common powers of understanding, a cold dessert of pureed fruit mixed with whipped cream or custard	idiot	A medieval jester
Step 3	foot	the terminal part of the vertebrate leg upon which an individual or item stands	end of lower limb	an adult foot
Step 4	fort	a strong or fortified place; *especially* **:** a fortified place occupied only by troops and surrounded with such works as a ditch, rampart, and parapet, a permanent army post -- often used in place names	soldiers' base	a child's toy fort
Step 5	fore	at an earlier time or period, in, toward, or adjacent to the front, golfing term	towards the front	Someone shouting 'fore' as they tee off on a golf course
Step 6	fire	the phenomenon of combustion manifested in light, flame, and heat, one of the four elements of the alchemists	burning	A blazing bonfire
Endword	five	number between four and six	square root of 25	A magnetic letter number '5'

GAME 2

GAME 2	target	Text Clue 1	Text Clue 2	Picture Clue
Starting word	lees			
Step 1	lies	to make an untrue statement with intent to deceive, to be or to stay at rest in a horizontal position or be prostrate	if you were having an affair, you might tell these to your partner	Baby lying on a mat kicking its legs
Step 2	dies	to pass from physical life	this is what is said to happen to a comic whose jokes are so bad that no-one laughs	Funeral procession with coffin
Step 3	dues	a payment or obligation required by law or custom	something that has to be paid	A bill
Step 4	duel	a formal combat with weapons fought between two persons in the presence of witnesses	often a fight to the death	Two men with swords ready to fight
Step 5	dull	mentally slow in perception or sensibility, low in saturation and low in lightness, lacking in force, intensity vivacity, resonance, or sharpness or brilliance	often used to describe British weather	Bored person
Step 6	pull	to exert force upon so as to cause or tend to cause motion toward the force	slang term for managing to attract someone of the opposite sex	Boy pulling garden wheelbarrow
Endword	poll	a questioning or canvassing of persons selected at random or by quota to obtain information or opinions to be analyzed	collection of votes	Voting box

GAME 3

GAME 3	target	Text Clue 1	Text Clue 2	Picture Clue
Starting word	Pity			
Step 1	pits	holes, shafts, or cavities in the ground: scooped-out places: areas sunken or depressed below the adjacent floor area: excavations in the earth from which mineral substances are taken	where mining done	Mining vista
Step 2	pins	small pointed pieces of wire with heads - used especially for fastening cloth	_ _ _ _ and needles	Pin cushion full of pins
Step 3	fins	an external membranous process of an aquatic animal (as a fish) used in propelling or guiding the bod	sharks have these	Shark's fin visible in the sea
Step 4	find	to come upon, often accidentally: to come upon by searching or effort: to discover by study or experiment, effort or management	to discover something	Game of hide and seek
Step 5	fond	having an affection or liking: foolishly tender: cherished with great affection	absence makes the heart grow _ _ _ _	Heart with an arrow through it
Step 6	food	material consisting essentially of protein, carbohydrate, and fat used in the body of an organism to sustain growth, repair, and vital processes and to furnish energy: nutriment in solid form: something that nourishes, sustains, or supplies	something you eat	Plate of dinner
Endword	good	of a favourable character or tendency: suitable or fit: commercially sound: agreeable, pleasant	opposite of bad	Girl making thumbs up sign

Pre and Post experiment questions

The following pages show how pre and post experiment sets of questions and scales included in this study were presented to participants. They are:

- Self reported demographic information e.g. data on age, culture. languages spoken, family socio-economic status, experience of and attitude to using computers;
- Communication style questionnaire (Stephen & Harrison, 1986) and communication style observation schedule to establish if habitual communication style is within norms, and if individual has a tendancy to aggressive, passive or assertive style (own checklist);
- Self Assessment Manikin (SAM) (Bradley & Lang, 1994) for self report of emotional valence and intensity;
- Positive and Negative Affect Scale (PANAS) for self report of affective state (Watson, Clark & Tellegen, 1988);
- Usability heuristics - set of 10 questions (based on Neilsen, 2005) for self report of usability of different conditions (and also included in order to give credence to the briefing given to participants that leads them to believe the focus is on usability of the system, rather than their emotional state;
- Affect intensity scale (short form) (Geuens M. & de Pelsmacker, 2002 for check on individual susceptibility to positive and negative affect;
- Big 5 personality test (Goldberg, 1992) to assess personality style;
- Set of questions relating to experience of the participant, and their views of whether they showed emotions, their attitude to the concept of affective machines, etc.(own questions).

1. Pre-experiment questions Subject ID……………Date……………….Time……………….

ABOUT YOU AND COMPUTERS

1 Are you male or female? Male Female

2 Please indicate your age.group

Under 20	21-25	26-35	36-45	46-55	56-65	66-75	75+

3 What is your cultural background/s ………………………………………………………

black / black brit			white			asian				mixed				other	
african	caribbean	other	british	irish	other	indian	pakistani	bangladeshi	other	white/caribbean	white/african	white/asian	other	chinese	other

What are the main cultural influences in your upbringing and current life?
....

4 What is your first / main language/s ……………………………….

5 What is your occupation? ………………………………………………………..

(or parents occupations if a student)……………………………………………………

If a student what subject are you studying? ……………………………………………..

6 What is your marital status?

single	married	cohabit	separated	divorced	widowed

7 Do you use a computer at home? Yes No

8 Do you use a computer at your place of work / study? Yes No

9 How long have you been using a computer for? …………………………

no experience	less than 1 yr	1-5 yrs	6-10 yrs	11-15yrs	16-20yrs	20+yrs

10 On average, how many hours a week do you use a computer? ....

none	less than 1hr	1-5 hrs	6-10 hrs	11-20 hrs	21-30 hrs	31-40 hrs	40+ hrs

11 What do you mainly use your computer for?

work	games	study	communication	hobby	other (Please specify)

12 How confident do you feel about using a computer?

extremely lacking in confidence	somewhat lacking in confidence	neither confident nor lacking confidence	somewhat confident	extremely confident

May I contact you for further research? If so please give name and contact details.
name……………………….email…………………………..phone……………………address……

Answer the following questionnaire to determine the style you use most when communicating. For each of the 42 items choose the word or phrase that best describes you in most situations and place a tick next to it. Answer all questions. Be honest with yourself. This isn't a test. There are no "right" or "wrong" answers, only those that apply best to you.

1.	Random		Patterned	
2.	Warm		Aloof	
3.	Spontaneous		Calculated	
4.	Less organised		More organised	
5.	Open		Reserved	
6.	Relationship orientated		Task orientated	
7.	Impulsive		Hesitant	
8.	Approachable		Distant	
9.	Risk		Safe play	
10.	Relaxed		Self-controlled	
11.	Unstructured		Structured	
12.	Sociable		Private	
13.	Flexible		Firm	
14.	Casual		Proper	
15.	Emotional		Logical	
16.	Easy going		Tense	
17.	Direct		Obtuse	
18.	Unfocused		Focused	
19.	Scattered		Disciplined	
20.	Less formal		More formal	
21.	Non-conformist		Conformist	
22.	Take charge		Go along	
23.	Boisterous		Calm	
24.	Challenging		Accepting	
25.	Overbearing		Shy	
26.	Extrovert		Introvert	
27.	Loud		Quiet	
28.	Initiate		Respond	
29.	Make statements		Ask questions	
30.	Outspoken		Withhold	
31.	Talkative		Listening	
32.	Pushy		Retiring	
33.	Lead		Follow	
34.	Overt		Covert	
35.	Argumentative		Agreeing	
36.	Unyielding		Compromise	
37.	Forward		Detached	
38.	Outgoing		Restrained	
39.	Take on		Avoid	
40.	Expanding		Not expanding	
41.	More dominant		Less dominant	
42.	Act		Wait	

Researcher ratings Subject ID……………Date………………Time………………

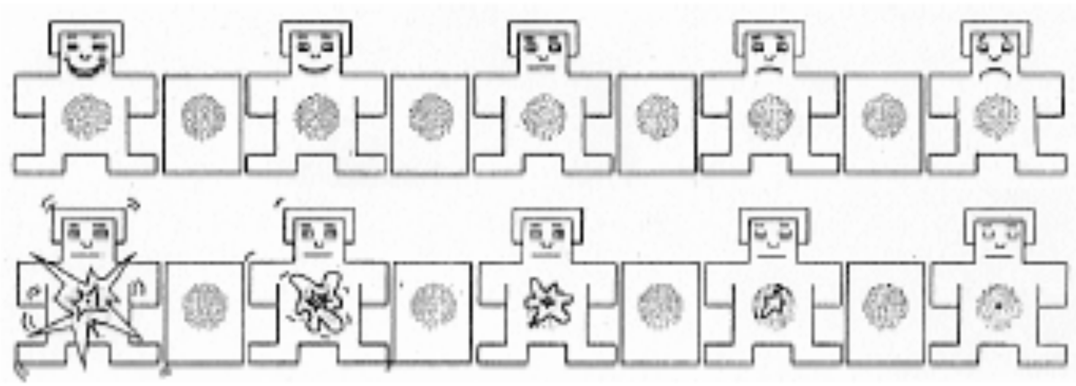
	very inapprop	rather inapprop	average	rather approp	very approp
EXPRESSION					
Overall expressive competency					
verbosity					
complexity					
pace					
cultural variation					
mixed messages					
interruptions					
stay on topic					
affirmations					
fillers					
monopolises / dominates					
irony/sarcasm					
warmth					
COMPREHENSION					
listening skills					
listener awareness					
turn taking					
verbal comprehension					
Non verbal comprehension / awareness					
active listening					
interruptions for clarification					
NVC					
NVC quantity					
NVC quality					

aggressive style	never				often
points					
frowns					
glares					
squints critically					
rigid posture					
loud critical voice					
fast clipped speech					
passive style					
sighs					
fidgets					
nods head often / pleading					
lack of facial animation					
smiles and nods a lot					
downcast eyes					
slumped posture					
low volume, meek					
up talk					
fast when anxious, slow + hesitant when doubtful					
assertive style					
open natural gestures					
attentive interested facial expression					
direct eye contact					
confident relaxed posture					
vocal volume appropriate expressive					
varied rate of speech					

Post – experiment questions

Subject ID. ………………….. Date / time…………………………..

HOW DO YOU FEEL NOW?



This scale consists of a number of words that describe different feelings and emotions. Read each item and then indicate to what extent you feel this way right now, that is at the present moment.

Rate how you feel right now.					
	not at all	a little	moderately	quite a bit	extremely
interested					
distressed					
excited					
upset					
strong					
guilty					
scared					
hostile					
enthusiastic					
proud					
irritable					
alert					
ashamed					
inspired					
nervous					
determined					
attentive					
jittery					
active					
afraid					

HOW USABLE WAS THE GAME?

As you were playing the game:

		not at all	a little	moderately	quite	a lot
1	Did you feel confident about where you were up to in the game and what to do next?					
2	Did the game seem to progress in a logical and natural way?					
3	Did you feel you had control during the game?					
4	Did the game seem to follow familiar rules and conventions?					
5	Did you encounter any system errors while playing the game?					
6	During the game, did you experience any difficulties in remembering what you had to do next?					
7	As you became more experienced in using the game, did you find you could speed up your interaction?					
8	Was the visual appearance of the game pleasing and uncluttered?					
9	Were you able to understand and deal with error messages efficiently?					
10	Did you receive enough explanation and help on how to play the game?					

Any other comments on usability?

……………………………………………………………………………………………

…………..

....

....

....

ABOUT YOU Indicate how you typically respond to the following events using the scale

		I never feel like that	I almost never feel like that	occasion-ally I feel like that	I usually feel like that	I almost always feel like that	I always feel like that
1	When I feel happy, it is a strong type of exuberance.						
2	I would characterise my happy moods as closer to contentment than joy.						
3	My happy moods are so strong that I feel like I'm in heaven.						
4	When I know I have done something very well, I feel relaxed and content rather than excited and elated.						
5	If I complete a task I thought was impossible, I am ecstatic.						
6	Sad movies deeply touch me.						
7	When I talk in front of a group for the first time, my voice gets shaky and my heart races.						
8	When I am feeling well, it's easy for me to go from being in a good mood to being really joyful.						
9	When I am happy, I feel like I'm bursting with joy						
10	When I'm happy, I feel very energetic						
11	When I am happy, the feeling is more like contentment and inner calm than one of exhilaration and excitement.						
12	When I do something wrong, I have strong feelings of shame and guilt.						
13	When I succeed at something, my reaction is calm and contented.						
14	When things are going good, I feel "on top of the world".						
15	When I do feel anxiety, it is normally very strong.						
16	When I feel happiness, it is a quiet type of contentment.						
17	When I'm happy, I bubble over with energy.						
18	When I feel guilty, this emotion is quite strong.						
19	When I'm nervous, I get shaky all over.						
20	When I'm happy, it's a feeling of being untroubled and content rather than being zestful and aroused.						

ABOUT YOU

On the following pages, there are phrases describing people's behaviours. Please use the rating scale below to describe how accurately each statement describes ***you***. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Please read each statement carefully, then tick a column to indicate how accurate the statement is about you.

	Very inaccurate	Moderately inaccurate	Neither	Moderately accurate	Very accurate
I am the life of the party.					
I feel little concern for others.					
I am always prepared.					
I get stressed out easily.					
I have a rich vocabulary.					
I don't talk a lot.					
I am interested in people.					
I leave my belongings around.					
I am relaxed most of the time.					
I have difficulty understanding abstract ideas.					
I feel comfortable around people.					
I insult people.					
I pay attention to details.					
I worry about things.					
I have a vivid imagination.					
I keep in the background.					
I sympathize with others' feelings.					
I make a mess of things.					
I seldom feel blue.					
I am not interested in abstract ideas.					
I start conversations.					
I'm not interested in other people's problems.					
I get chores done right away.					
I am easily disturbed.					
I have excellent ideas.					
I have little to say.					
I have a soft heart.					
I often forget to put things back in their proper place.					
I get upset easily.					
I do not have a good imagination.					
I talk to a lot of different people at parties.					
I am not really interested in others.					
I like order.					
I change my mood a lot.					
I am quick to understand things.					
I don't like to draw attention to myself.					
I take time out for others.					
I shirk my duties.					
I have frequent mood swings.					
I use difficult words.					
I don't mind being the center of attention.					
I feel others' emotions.					
I follow a schedule.					
I get irritated easily.					
I spend time reflecting on things.					
I am quiet around strangers.					
I make people feel at ease.					
I am exacting in my work.					
I often feel blue.					
I am full of ideas.					

DURING THE GAME

1. During the ‘word ladder’ game, did you feel any emotions? Yes / No
If so what did you feel?
……………………………………………………………………………………………………

……………………………………………………………………………………………………

2. If you felt any emotions, do you think you showed them in any way? Yes / No
If so how did you show them?
……………………………………………………………………………………………………

……………………………………………………………………………………………………

3. Did the word ladder game seem to adapt itself to you in any way? Yes / No
If so how did it adapt?
……………………………………………………………………………………………………

……………………………………………………………………………………………………

4. Do you believe that the computer recognised any of your emotions? Yes / No
If so what emotions did it recognise and how do you think it recognised them?
……………………………………………………………………………………………………

……………………………………………………………………………………………………

5. Do you believe that the computer responded to any of your emotions? Yes / No
If so how did it respond?
……………………………………………………………………………………………………

……………………………………………………………………………………………………

6. In the future all sorts of computer systems and other electronic devices could be designed to recognise your emotions very efficiently and respond to them in different ways, such as by changing the speed of presentation or altering the level of difficulty or altering the type of information presented. What are your views on this?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

7. Like most new developments, emotional computer systems could greatly improve services, for example: e-learning courses. They could also be used for commercial and other purposes, for example to target advertising to your mood. Overall would you like emotional recognition systems to be developed or not?

Feel strongly - they should never be developed	Feel mildly - on the whole better if not developed	Don’t care if developed or not	Feel mildly - best if developed	Feel strongly - they should be developed

Appendix F. Quantitative Analysis

These tables summarise the significant results and trends found from analysis in SPSS in the first experiment.

General Linear Model results follow to indicate significant at .05

VARIABLE	SCORING	SIG	IN ACTED AFFECTIVE CONDITION
believed	Yes/no	.003	More participants believed the system was affective
believed game adapted	Yes/no	.005	More believed game adapted to them
rungs completed	Count	.007	More rungs of game completed
SAM valence	Rating Score	.002	After playing game more reported themselves as feeling happier
observed communication style	Rating score	.048	I observed more assertive – aggressive communication style.
Observed valance	Rating score	.001	Lower no. = more positive valance

VARIABLE	SCORING	SIG	IN TOLD AFFECTIVE CONDITION
Blinkrate	count	.037	Participants blinked more
Self report aggressive comm'n style	Rating score	.030	More likely to report themselves as generally having an aggressive communication style
AIM negative score	Test score	.043	More likely to report themselves as generally being susceptible to negative affect
Self reported showed emotion	Yes / no	.042	More likely to say they had showed their emotions
Observed aggressive communication style	Rating score	.022	More often observed to use aggressive communication style
Observed assertive communication style	Rating score	.023	More often observed to use assertive communication style
More Clues	count	.018	Less likely to ask for clues
Observed intensity	Rating score	.034	Lower no. = higher arousal

This table summarises trends in data - non significant results found from analysis in SPSS in the first experiment.

VARIABLE	SCORING	SIG	IN ACTED AFFECTIVE CONDITION
reported PC use for communication	Yes/no	.064	More likely to use PC at home for communication
reported PC use for games	Yes/no	.072	More likely to use PC at home for playing games

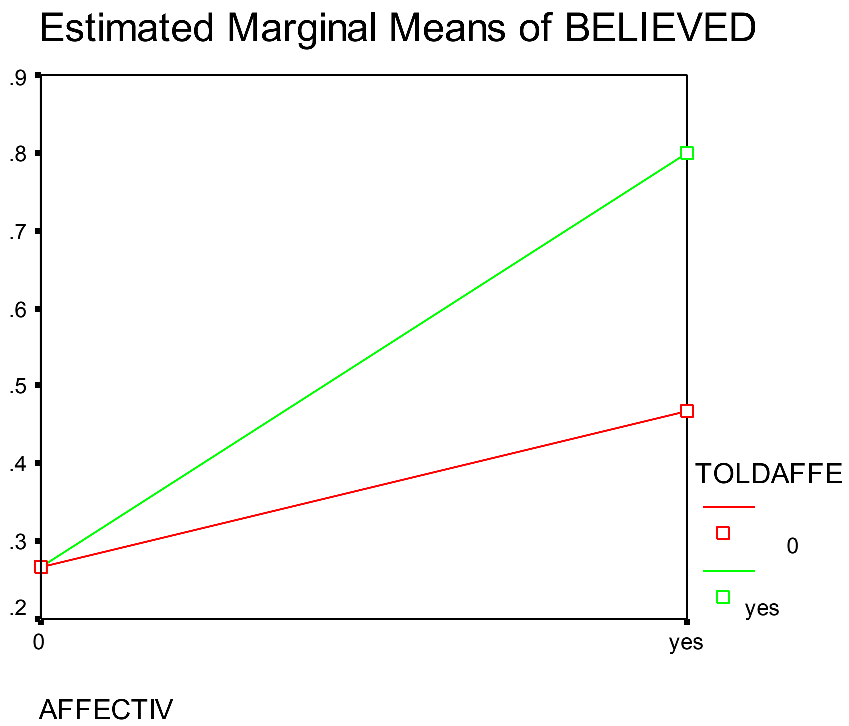
VARIABLE	SCORING	SIG	IN TOLD AFFECTIVE CONDITION
reported confidence in computer use	Rating score	.065	Less likely to report themselves as confident PC users
Vocalics	count	.080	More vocal during game interaction if told
SAM arousal	Self report rating score	.068	Reported selves as less aroused if told

Tests of Between-Subjects Effects

Dependent Variable: BELIEVED in affective system

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.850(a)	3	.950	4.433	.007
Intercept	12.150	1	12.150	56.700	.000
TOLDAFFE	.417	1	.417	1.944	.169
AFFECTIV	2.017	1	2.017	9.411	.003
TOLDAFFE * AFFECTIV	.417	1	.417	1.944	.169
Error	12.000	56	.214		
Total	27.000	60			
Corrected Total	14.850	59			

a R Squared = .192 (Adjusted R Squared = .149)



Report

BELIEVED

GROUP	Mean	N	Std. Deviation
told+affective	.80	15	.414
told + not affective	.27	15	.458
not told + affective	.47	15	.516
not told + not affective	.27	15	.458
Total	.45	60	.502

	Acted affective	Standard
Told affective	mean .80 SD .414	mean .27 SD .485
Not told affective	mean .47 SD .516	mean .27 SD .458

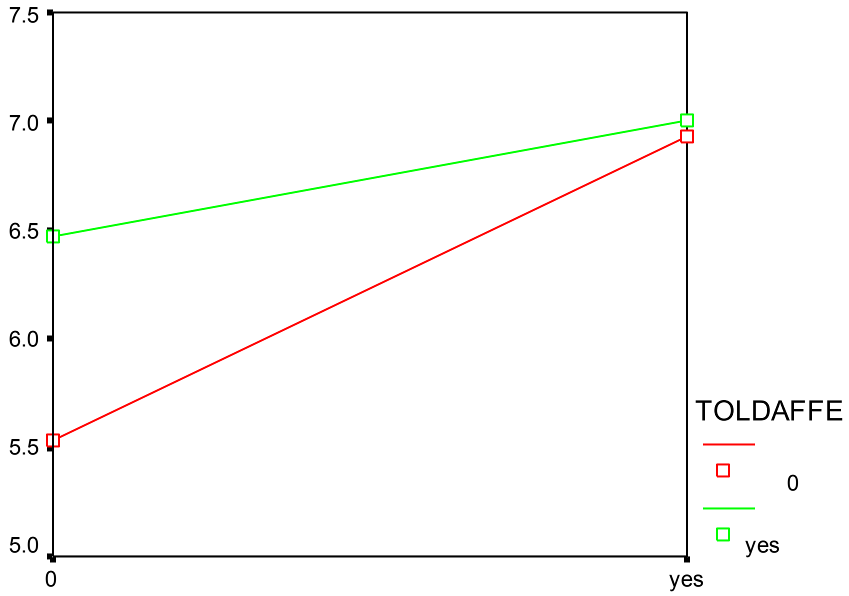
Tests of Between-Subjects Effects

Dependent Variable: RUNGS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	20.583(a)	3	6.861	3.827	.015
Intercept	2522.017	1	2522.017	1406.703	.000
TOLDAFFE	3.750	1	3.750	2.092	.154
AFFECTIV	14.017	1	14.017	7.818	.007
TOLDAFFE * AFFECTIV	2.817	1	2.817	1.571	.215
Error	100.400	56	1.793		
Total	2643.000	60			
Corrected Total	120.983	59			

a R Squared = .170 (Adjusted R Squared = .126)

Estimated Marginal Means of RUNGS



Report

RUNGS

GROUP	Mean	N	Std. Deviation
told+affective	7.00	15	.000
told + not affective	6.47	15	1.246
not told + affective	6.93	15	.258
not told + not affective	5.53	15	2.356
Total	6.48	60	1.432

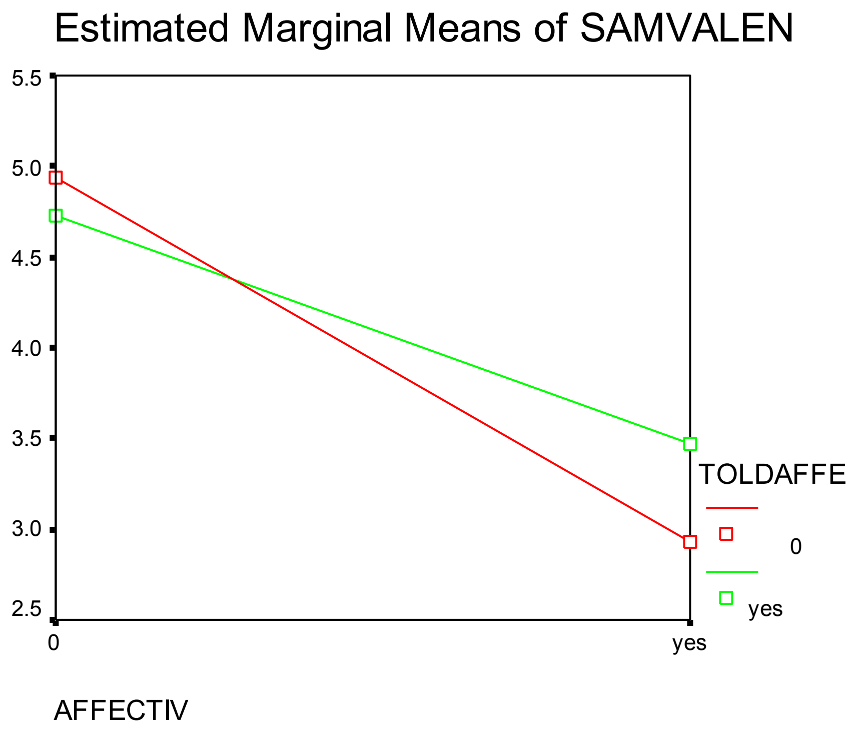
	Acted affective	Standard
Told affective	mean 7.00 SD .000	mean 6.47 SD 1.246
Not told affective	mean 6.93 SD .258	mean 5.53 SD 2.356

Tests of Between-Subjects Effects

Dependent Variable: SAMVALEN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	42.450(a)	3	14.150	3.626	.018
Intercept	968.017	1	968.017	248.058	.000
TOLDAFFE	.417	1	.417	.107	.745
AFFECTIV	40.017	1	40.017	10.254	.002
TOLDAFFE * AFFECTIV	2.017	1	2.017	.517	.475
Error	218.533	56	3.902		
Total	1229.000	60			
Corrected Total	260.983	59			

a R Squared = .163 (Adjusted R Squared = .118)



Report

SAMVALEN

GROUP	Mean	N	Std. Deviation
told+affective	3.47	15	1.807
told + not affective	4.73	15	1.831
not told + affective	2.93	15	1.792
not told + not affective	4.93	15	2.404
Total	4.02	60	2.103

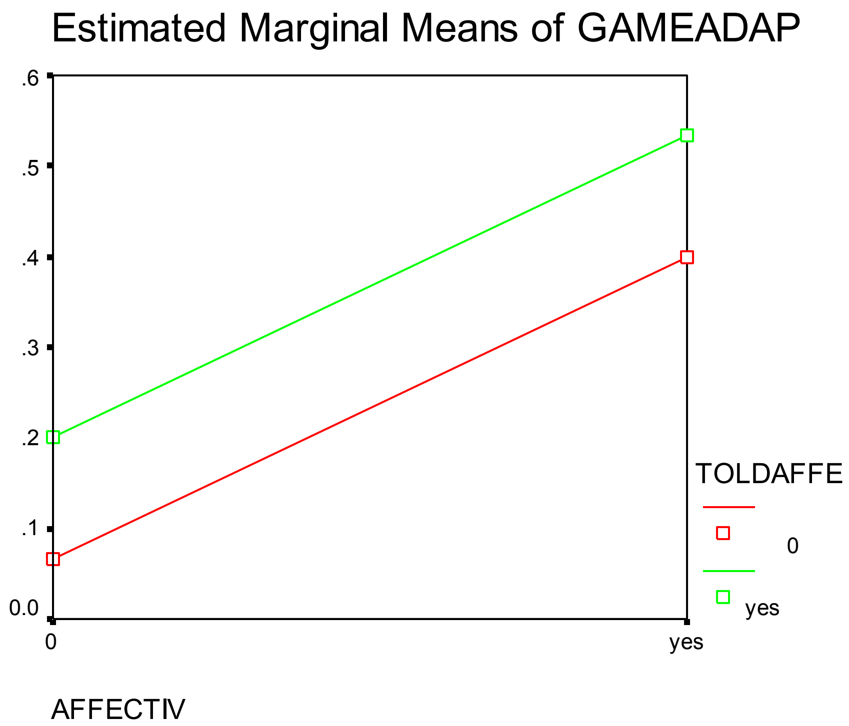
	Acted affective		Standard	
Told affective	mean	3.47	mean	4.73
	SD	1.807	SD	1.831
Not told affective	mean	2.93	mean	4.93
	SD	1.792	SD	2.404

Tests of Between-Subjects Effects

Dependent Variable: GAMEADAP

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.933(a)	3	.644	3.383	.024
Intercept	5.400	1	5.400	28.350	.000
TOLDAFFE	.267	1	.267	1.400	.242
AFFECTIV	1.667	1	1.667	8.750	.005
TOLDAFFE * AFFECTIV	.000	1	.000	.000	1.000
Error	10.667	56	.190		
Total	18.000	60			
Corrected Total	12.600	59			

a R Squared = .153 (Adjusted R Squared = .108)



Report

GAMEADAP

GROUP	Mean	N	Std. Deviation
told+affective	.53	15	.516
told + not affective	.20	15	.414
not told + affective	.40	15	.507
not told + not affective	.07	15	.258
Total	.30	60	.462

	Acted affective	Standard
Told affective	mean .53 SD .516	mean .20 SD .414
Not told affective	mean .40 SD .507	mean .07 SD .258

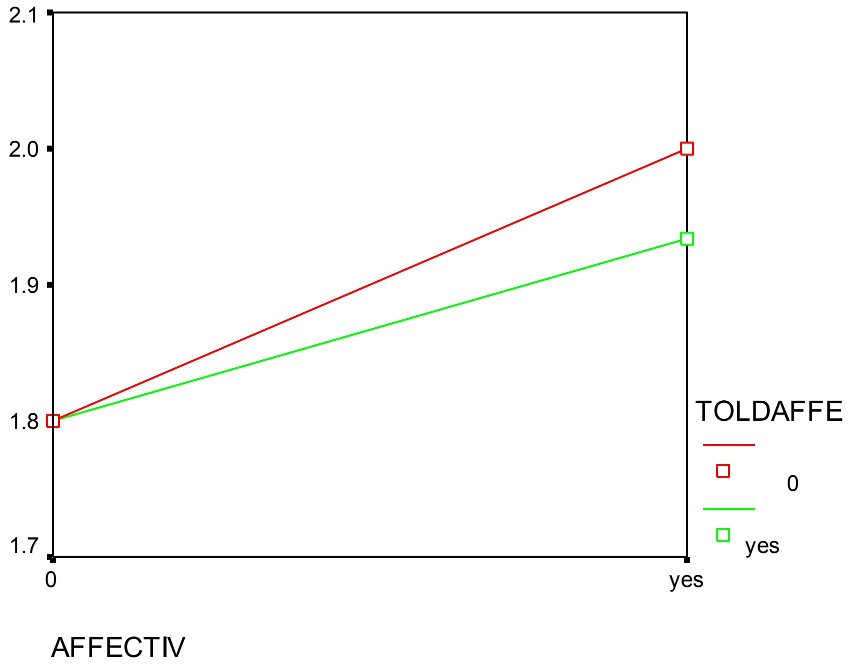
Tests of Between-Subjects Effects

Dependent Variable: OBSCOMST

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.450(a)	3	.150	1.465	.234
Intercept	212.817	1	212.817	2078.674	.000
TOLDAFFE	.017	1	.017	.163	.688
AFFECTIV	.417	1	.417	4.070	.048
TOLDAFFE * AFFECTIV	.017	1	.017	.163	.688
Error	5.733	56	.102		
Total	219.000	60			
Corrected Total	6.183	59			

a R Squared = .073 (Adjusted R Squared = .023)

Estimated Marginal Means of OBSCOMST



Report

OBSCOMST

GROUP	Mean	N	Std. Deviation
told+affective	1.93	15	.258
told + not affective	1.80	15	.414
not told + affective	2.00	15	.000
not told + not affective	1.80	15	.414
Total	1.88	60	.324

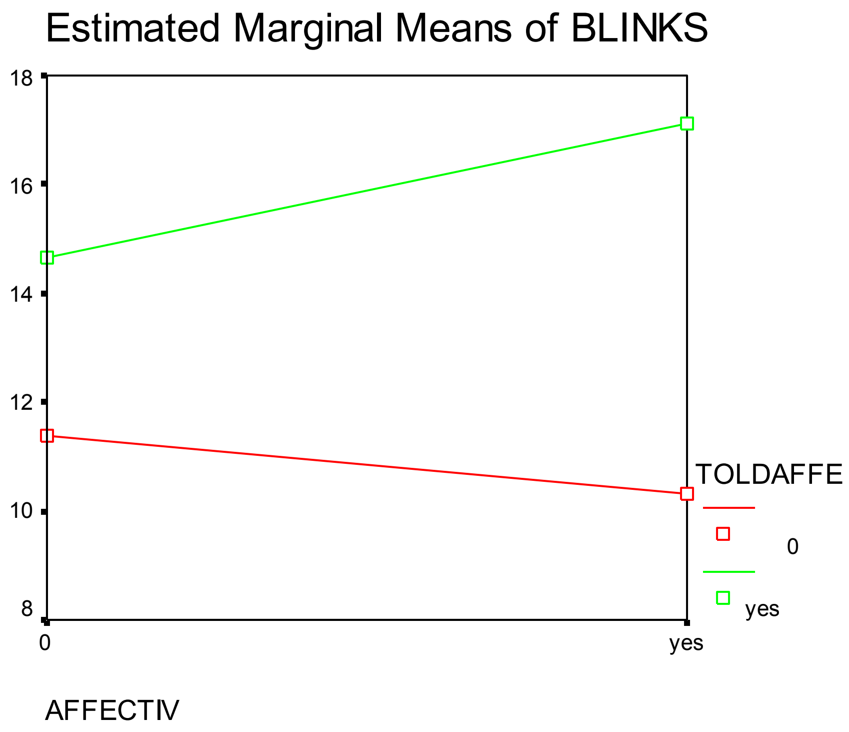
	Acted affective		Standard	
Told affective	mean	1.93	mean	1.80
	SD	.258	SD	.414
Not told affective	mean	2.00	mean	1.80
	SD	.000	SD	.414

Tests of Between-Subjects Effects

Dependent Variable: BLINKS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	434.183(a)	3	144.728	1.741	.169
Intercept	10746.817	1	10746.817	129.313	.000
TOLDAFFE	380.017	1	380.017	4.573	.037
AFFECTIV	7.350	1	7.350	.088	.767
TOLDAFFE * AFFECTIV	46.817	1	46.817	.563	.456
Error	4654.000	56	83.107		
Total	15835.000	60			
Corrected Total	5088.183	59			

a R Squared = .085 (Adjusted R Squared = .036)



Report

BLINKS

GROUP	Mean	N	Std. Deviation
told+affective	17.13	15	9.094
told + not affective	14.67	15	10.991
not told + affective	10.33	15	9.076
not told + not affective	11.40	15	6.822
Total	13.38	60	9.287

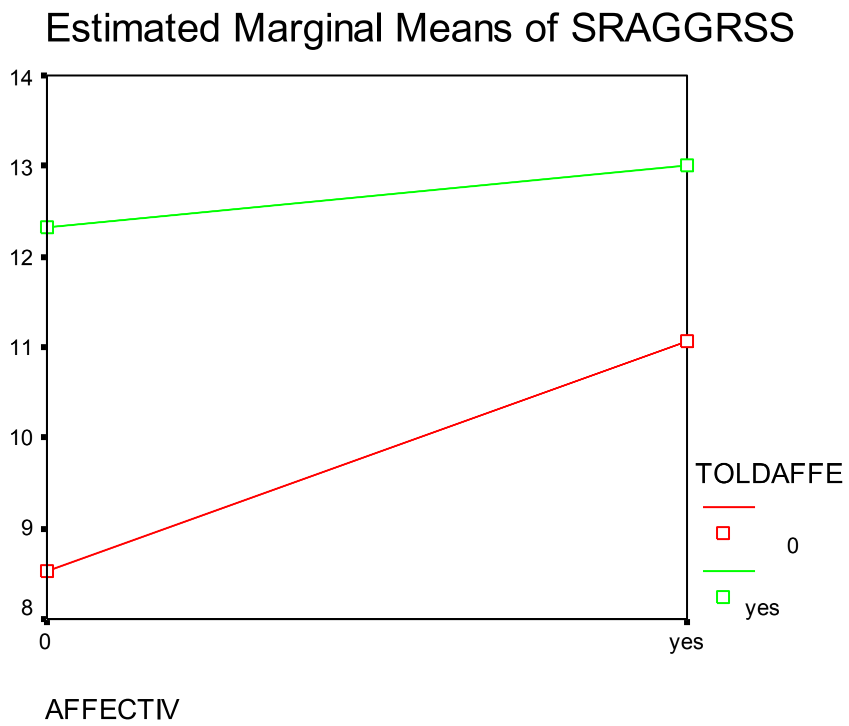
	Acted affective		Standard	
Told affective	mean	17.13	mean	14.67
	SD	9.094	SD	10.991
Not told affective	mean	10.33	mean	11.40
	SD	9.076	SD	6.822

Tests of Between-Subjects Effects

Dependent Variable: SRAGGRSS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	174.733(a)	3	58.244	2.336	.083
Intercept	7571.267	1	7571.267	303.718	.000
TOLDAFFE	123.267	1	123.267	4.945	.030
AFFECTIV	38.400	1	38.400	1.540	.220
TOLDAFFE * AFFECTIV	13.067	1	13.067	.524	.472
Error	1396.000	56	24.929		
Total	9142.000	60			
Corrected Total	1570.733	59			

a R Squared = .111 (Adjusted R Squared = .064)



Report

SRAGGRSS

GROUP	Mean	N	Std. Deviation
told+affective	13.00	15	4.209
told + not affective	12.33	15	4.639
not told + affective	11.07	15	5.922
not told + not affective	8.53	15	5.041
Total	11.23	60	5.160

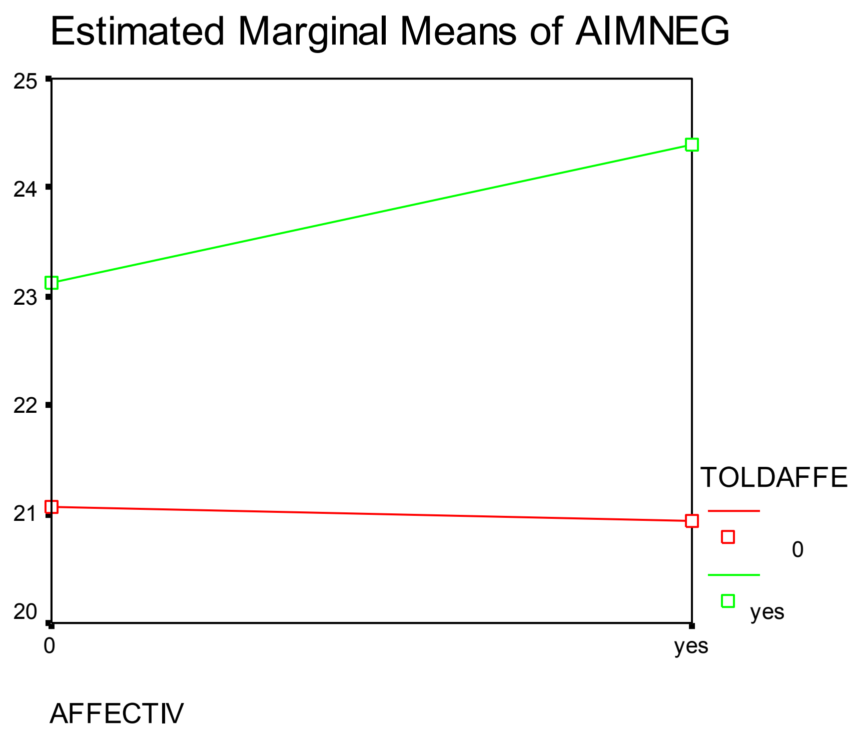
	Acted affective		Standard	
Told affective	mean	13.00	mean	12.33
	SD	4.209	SD	4.639
Not told affective	mean	11.07	mean	8.53
	SD	5.922	SD	5.041

Tests of Between-Subjects Effects

Dependent Variable: AIMNEG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	125.042(a)	3	41.681	1.595	.201
Intercept	29536.576	1	29536.576	1130.335	.000
TOLDAFFE	112.608	1	112.608	4.309	.043
AFFECTIV	4.692	1	4.692	.180	.673
TOLDAFFE * AFFECTIV	7.270	1	7.270	.278	.600
Error	1437.195	55	26.131		
Total	31184.000	59			
Corrected Total	1562.237	58			

a R Squared = .080 (Adjusted R Squared = .030)



Report

AIMNEG

GROUP	Mean	N	Std. Deviation
told+affective	24.40	15	5.262
told + not affective	23.13	15	5.370
not told + affective	20.93	15	4.743
not told + not affective	21.07	14	5.045
Total	22.41	59	5.190

	Acted affective		Standard	
Told affective	mean	24.40	mean	23.13
	SD	5.262	SD	5.370
Not told affective	mean	20.93	mean	21.07
	SD	4.743	SD	5.045

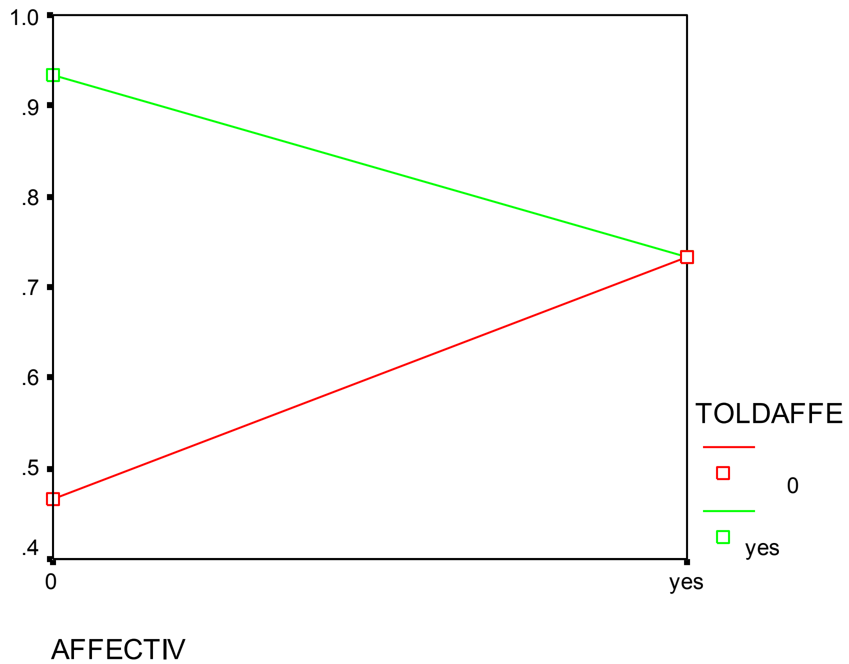
Tests of Between-Subjects Effects

Dependent Variable: SHOWEMOT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.650(a)	3	.550	2.924	.042
Intercept	30.817	1	30.817	163.835	.000
TOLDAFFE	.817	1	.817	4.342	.042
AFFECTIV	.017	1	.017	.089	.767
TOLDAFFE * AFFECTIV	.817	1	.817	4.342	.042
Error	10.533	56	.188		
Total	43.000	60			
Corrected Total	12.183	59			

a R Squared = .135 (Adjusted R Squared = .089)

Estimated Marginal Means of SHOWEMOT



Report

SHOWEMOT

GROUP	Mean	N	Std. Deviation
told+affective	.73	15	.458
told + not affective	.93	15	.258
not told + affective	.73	15	.458
not told + not affective	.47	15	.516
Total	.72	60	.454

Report

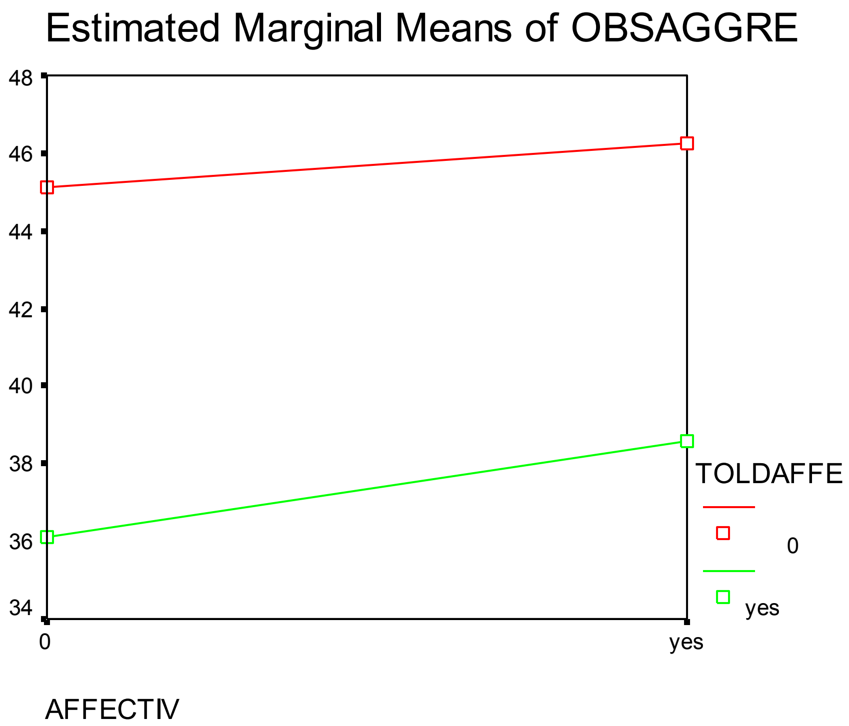
	Acted affective		Standard	
Told affective	mean	.73	mean	.93
	SD	.458	SD	.258
Not told affective	mean	.73	mean	.47
	SD	.458	SD	.516

Tests of Between-Subjects Effects

Dependent Variable: OBSAGGRE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1096.933(a)	3	365.644	1.938	.134
Intercept	103501.067	1	103501.067	548.661	.000
TOLDAFFE	1041.667	1	1041.667	5.522	.022
AFFECTIV	48.600	1	48.600	.258	.614
TOLDAFFE * AFFECTIV	6.667	1	6.667	.035	.852
Error	10564.000	56	188.643		
Total	115162.000	60			
Corrected Total	11660.933	59			

a R Squared = .094 (Adjusted R Squared = .046)



Report

OBSAGGRE

GROUP	Mean	N	Std. Deviation
told+affective	38.60	15	13.569
told + not affective	36.13	15	13.948
not told + affective	46.27	15	12.775
not told + not affective	45.13	15	14.584
Total	41.53	60	14.059

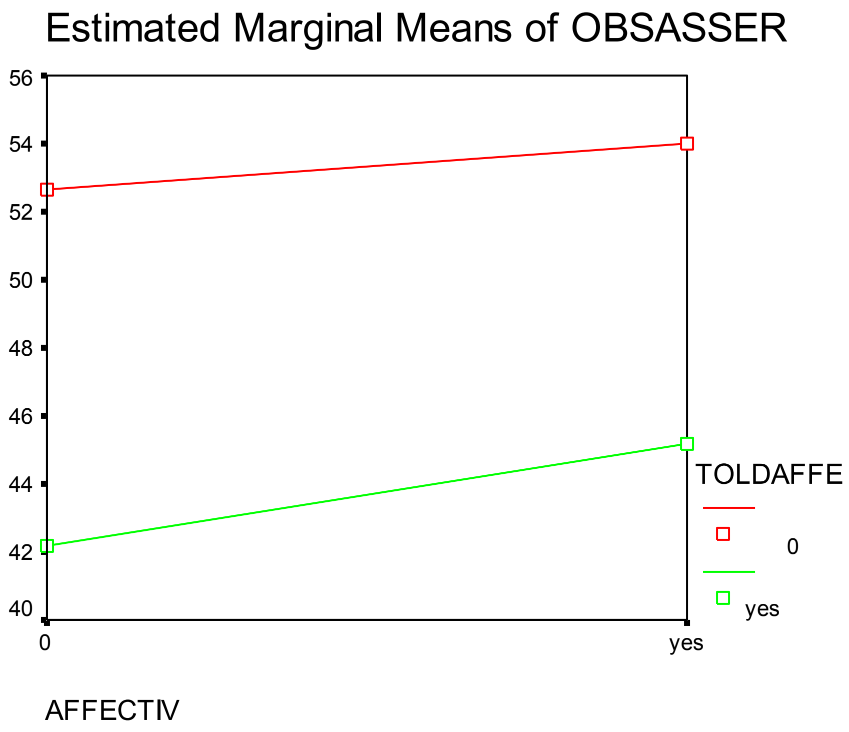
	Acted affective		Standard	
Told affective	mean	38.60	mean	36.13
	SD	13.569	SD	13.948
Not told affective	mean	46.27	mean	45.13
	SD	12.775	SD	14.584

Tests of Between-Subjects Effects

Dependent Variable: OBSASSER

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1472.850(a)	3	490.950	1.918	.137
Intercept	141232.017	1	141232.017	551.682	.000
TOLDAFFE	1392.017	1	1392.017	5.438	.023
AFFECTIV	70.417	1	70.417	.275	.602
TOLDAFFE * AFFECTIV	10.417	1	10.417	.041	.841
Error	14336.133	56	256.002		
Total	157041.000	60			
Corrected Total	15808.983	59			

a R Squared = .093 (Adjusted R Squared = .045)



Report

OBSASSER

GROUP	Mean	N	Std. Deviation
told+affective	45.20	15	15.826
told + not affective	42.20	15	16.280
not told + affective	54.00	15	14.914
not told + not affective	52.67	15	16.914
Total	48.52	60	16.369

	Acted affective		Standard	
Told affective	mean	45.20	mean	42.20
	SD	15.826	SD	16.280
Not told affective	mean	54.00	mean	52.67
	SD	14.914	SD	16.914

Observed Valency

Between-Subjects Factors

		Value Label	N
toldaffe	0		30
	1	yes	30
affectiv	0		30
	1	yes	30

Tests of Between-Subjects Effects

Dependent Variable:av.val

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11.015[a]	3	3.672	10.631	.000
Intercept	1796.348	1	1796.348	5201.210	.000
toldaffe	2.596	1	2.596	7.516	.008
affectiv	4.363	1	4.363	12.633	.001
toldaffe * affectiv	4.056	1	4.056	11.744	.001
Error	19.341	56	.345		
Total	1826.704	60			
Corrected Total	30.356	59			

a. R Squared = .363 (Adjusted R Squared = .329)

Report

av.val

group	Mean	N	Std. Deviation
1	4.7340	15	.37571
2	5.7933	15	.75059
3	5.6700	15	.53874
4	5.6893	15	.62186
Total	5.4717	60	.71729

Observed Intensities

Between-Subjects Factors

		Value Label	N
toldaffe	0		30
	1	yes	30
affectiv	0		30
	1	yes	30

Tests of Between-Subjects Effects

Dependent Variable:av.int

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.703[a]	3	.901	1.690	.180
Intercept	1149.050	1	1149.050	2155.226	.000
toldaffe	2.526	1	2.526	4.737	.034
affectiv	.176	1	.176	.330	.568
toldaffe * affectiv	.002	1	.002	.003	.956
Error	29.856	56	.533		
Total	1181.610	60			
Corrected Total	32.559	59			

a. R Squared = .083 (Adjusted R Squared = .034)

Report

av.int

group	Mean	N	Std. Deviation
1	4.1220	15	.42007
2	4.2200	15	.78793
3	4.5220	15	.89474
4	4.6407	15	.73126
Total	4.3762	60	.74287

Clues requested

Report

reqclue

group	Mean	N	Std. Deviation
told+affective	1.80	15	1.082
told + not affective	1.53	15	1.407
not told + affective	2.40	15	1.242
not told + not affective	2.47	15	1.125
Total	2.05	60	1.254

Between-Subjects Factors

		Value Label	N
toldaffe	0		30
	1	yes	30
affectiv	0		30
	1	yes	30

Tests of Between-Subjects Effects

Dependent Variable:reqclue

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	9.383[a]	3	3.128	2.099	.111
Intercept	252.150	1	252.150	169.174	.000
toldaffe	8.817	1	8.817	5.915	.018
affectiv	.150	1	.150	.101	.752
toldaffe * affectiv	.417	1	.417	.280	.599
Error	83.467	56	1.490		
Total	345.000	60			
Corrected Total	92.850	59			

a. R Squared = .101 (Adjusted R Squared = .053)

POST HOC INDEPENDENT SAMPLES T TEST RESULTS

SUMMARY

RUNGS T test significance between groups 1+2, 1+4, 2+3, 2+4, 1+4

Groups	1	2	3	4
1		.001	.040	.000
2			.004	.008
3				.000
4				

SELF ASSESSMENT MANIKIN VALENCE no significance

Groups	1	2	3	4
1		.560	.718	.034
2			.803	.055
3				.039
4				

BLINKS no significance

Groups	1	2	3	4
1		.605	.676	.150
2			.405	.089
3				.393
4				

AIM NEGATIVE SCORE no significance

Groups	1	2	3	4
1		.855	.940	.618
2			.779	.767
3				.512
4				

BELIEVED THEY HAD SHOWED EMOTIONS significance between groups 1+2, 3+2, 3+4

Groups	1	2	3	4
1		.002	1.000	.067
2			.002	.000
3				.067
4				

RUNGS T test significance between groups 1+2, 1+4, 2+3, 2+4, 1+4

Groups	1	2	3	4
1		.001	.040	.000
2			.004	.008
3				.000
4				

groups 1 and 2 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
RUNGS	told+affective	15	7.00	.000	.000
	told + not affective	15	6.47	1.246	.322

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
RUNGS	Equal variances assumed	14.145	.001	1.658	28	.109	.53	.322	-.126	1.192
	Equal variances not assumed			1.658	14.000	.120	.53	.322	-.157	1.223

groups 1 + 3 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
RUNGS	told+affective	15	7.00	.000	.000
	not told + affective	15	6.93	.258	.067

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
RUNGS	Equal variances assumed	4.639	.040	1.000	28	.326	.07	.067	-.070	.203
	Equal variances not assumed			1.000	14.000	.334	.07	.067	-.076	.210

groups 1 + 4

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
RUNGS	told+affective	15	7.00	.000	.000
	not told + not affective	15	5.53	2.356	.608

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RUNGS	Equal variances assumed	39.424	.000	2.411	28	.023	1.47	.608	.220	2.713
	Equal variances not assumed			2.411	14.000	.030	1.47	.608	.162	2.772

groups 2 + 3 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
RUNGS	told + not affective	15	6.47	1.246	.322
	not told + affective	15	6.93	.258	.067

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RUNGS	Equal variances assumed	9.692	.004	-1.420	28	.167	-.47	.329	-1.140	.206
	Equal variances not assumed			-1.420	15.200	.176	-.47	.329	-1.166	.233

groups 2 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
RUNGS	told + not affective	15	6.47	1.246	.322
	not told + not affective	15	5.53	2.356	.608

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RUNGS	Equal variances assumed	8.182	.008	1.356	28	.186	.93	.688	-.476	2.343
	Equal variances not assumed			1.356	21.261	.189	.93	.688	-.497	2.364

groups 3 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
RUNGS	not told + affective	15	6.93	.258	.067
	not told + not affective	15	5.53	2.356	.608

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RUNGS	Equal variances assumed	33.416	.000	2.287	28	.030	1.40	.612	.146	2.654
	Equal variances not assumed			2.287	14.336	.038	1.40	.612	.090	2.710

SELF ASSESSMENT MANIKIN VALENCE no significance

Groups	1	2	3	4
1		.560	.718	.034
2			.803	.055
3				.039
4				

groups 1 + 2

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SAMVALEN	told+affective	15	3.47	1.807	.467
	told + not affective	15	4.73	1.831	.473

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SAMVALEN	Equal variances assumed	.348	.560	-1.907	28	.067	-1.27	.664	-2.627	.094
	Equal variances not assumed			-1.907	27.995	.067	-1.27	.664	-2.627	.094

groups 1 + 3 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SAMVALEN	told+affective	15	3.47	1.807	.467
	not told + affective	15	2.93	1.792	.463

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
SAMVALEN	Equal variances assumed	.133	.718	.812	28	.424	.53	.657	-.813	1.879
	Equal variances not assumed			.812	27.998	.424	.53	.657	-.813	1.879

groups 1 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SAMVALEN	told+affective	15	3.47	1.807	.467
	not told + not affective	15	4.93	2.404	.621

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SAMVALEN	Equal variances assumed	4.988	.034	-1.888	28	.069	-1.47	.777	-3.058	.124
	Equal variances not assumed			-1.888	25.993	.070	-1.47	.777	-3.063	.130

groups 2 + 3 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SAMVALEN	told + not affective	15	4.73	1.831	.473
	not told + affective	15	2.93	1.792	.463

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SAMVALEN	Equal variances assumed	.063	.803	2.721	28	.011	1.80	.661	.445	3.155
	Equal variances not assumed			2.721	27.987	.011	1.80	.661	.445	3.155

groups 2 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SAMVALEN	told + not affective	15	4.73	1.831	.473
	not told + not affective	15	4.93	2.404	.621

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SAMVALEN	Equal variances assumed	4.025	.055	-.256	28	.800	-.20	.780	-1.798	1.398
	Equal variances not assumed			-.256	26.151	.800	-.20	.780	-1.804	1.404

groups 3 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SAMVALEN	not told + affective	15	2.93	1.792	.463
	not told + not affective	15	4.93	2.404	.621

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SAMVALEN	Equal variances assumed	4.710	.039	-2.583	28	.015	-2.00	.774	-3.586	-.414
	Equal variances not assumed			-2.583	25.883	.016	-2.00	.774	-3.592	-.408

BLINKS no significance

Groups	1	2	3	4
1		.605	.676	.150
2			.405	.089
3				.393
4				

groups 1 + 2 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
BLINKS	told+affective	15	17.13	9.094	2.348
	told + not affective	15	14.67	10.991	2.838

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
BLINKS	Equal variances assumed	.273	.605	.670	28	.509	2.47	3.683	-5.078	10.012
	Equal variances not assumed			.670	27.051	.509	2.47	3.683	-5.090	10.024

groups 1 + 3 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
BLINKS	told+affective	15	17.13	9.094	2.348
	not told + affective	15	10.33	9.076	2.344

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
BLINKS	Equal variances assumed	.179	.676	2.050	28	.050	6.80	3.317	.005	13.595
	Equal variances not assumed			2.050	28.000	.050	6.80	3.317	.005	13.595

groups 1 + 4

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
BLINKS	told+affective	15	17.13	9.094	2.348
	not told + not affective	15	11.40	6.822	1.761

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
BLINKS	Equal variances assumed	2.189	.150	1.953	28	.061	5.73	2.935	-.279	11.746
	Equal variances not assumed			1.953	25.968	.062	5.73	2.935	-.301	11.767

groups 2 + 3 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
BLINKS	told + not affective	15	14.67	10.991	2.838
	not told + affective	15	10.33	9.076	2.344

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
BLINKS	Equal variances assumed	.714	.405	1.177	28	.249	4.33	3.680	-3.206	11.872
	Equal variances not assumed			1.177	27.033	.249	4.33	3.680	-3.218	11.885

groups 2 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
BLINKS	told + not affective	15	14.67	10.991	2.838
	not told + not affective	15	11.40	6.822	1.761

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
BLINKS	Equal variances assumed	3.094	.089	.978	28	.336	3.27	3.340	-3.575	10.109
	Equal variances not assumed			.978	23.393	.338	3.27	3.340	-3.637	10.170

groups 3 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
BLINKS	not told + affective	15	10.33	9.076	2.344
	not told + not affective	15	11.40	6.822	1.761

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
BLINKS	Equal variances assumed	.754	.393	-.364	28	.719	-1.07	2.932	-7.072	4.939
	Equal variances not assumed			-.364	25.992	.719	-1.07	2.932	-7.093	4.960

AIM NEGATIVE SCORE no significance

Groups	1	2	3	4
1		.855	.940	.618
2			.779	.767
3				.512
4				

groups 1 + 2 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
AIMNEG	told+affective	15	24.40	5.262	1.359
	told + not affective	15	23.13	5.370	1.387

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
AIM NEG	Equal variances assumed	.034	.855	.653	28	.519	1.27	1.941	-2.710	5.243
	Equal variances not assumed			.653	27.988	.519	1.27	1.941	-2.710	5.243

groups 1 + 3 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
AIMNEG	told+affective	15	24.40	5.262	1.359
	not told + affective	15	20.93	4.743	1.225

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
AIM NEG	Equal variances assumed	.006	.940	1.895	28	.068	3.47	1.829	-.280	7.213
	Equal variances not assumed			1.895	27.704	.069	3.47	1.829	-.282	7.215

groups 1 + 4

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
AIMNEG	told+affective	15	24.40	5.262	1.359
	not told + not affective	14	21.07	5.045	1.348

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
AIMNEG	Equal variances assumed	.254	.618	1.736	27	.094	3.33	1.917	-.605	7.262
	Equal variances not assumed			1.739	26.976	.093	3.33	1.914	-.599	7.256

groups 2 + 3 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
AIMNEG	told + not affective	15	23.13	5.370	1.387
	not told + affective	15	20.93	4.743	1.225

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
AIMNEG	Equal variances assumed	.081	.779	1.189	28	.244	2.20	1.850	-1.589	5.989
	Equal variances not assumed			1.189	27.579	.244	2.20	1.850	-1.592	5.992

groups 2 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
AIMNEG	told + not affective	15	23.13	5.370	1.387
	not told + not affective	14	21.07	5.045	1.348

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
AIMNEG	Equal variances assumed	.090	.767	1.064	27	.297	2.06	1.938	-1.915	6.039
	Equal variances not assumed			1.066	26.998	.296	2.06	1.934	-1.907	6.030

groups 3+ 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
AIMNEG	not told + affective	15	20.93	4.743	1.225
	not told + not affective	14	21.07	5.045	1.348

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
AIMNEG	Equal variances assumed	.442	.512	-.076	27	.940	-.14	1.818	-3.867	3.591
	Equal variances not assumed			-.076	26.530	.940	-.14	1.822	-3.879	3.602

BELIEVED THEY HAD SHOWED EMOTIONS significance between groups 1+2, 3+2, 3+4

Groups	1	2	3	4
1		.002	1.000	.067
2			.002	.000
3				.067
4				

groups 1 + 2

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SHOWEMOT	told+affective	15	.73	.458	.118
	told + not affective	15	.93	.258	.067

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
SHOW EMOT	Equal variances assumed	11.146	.002	-1.474	28	.152	-.20	.136	-.478	.078
	Equal variances not assumed			-1.474	22.090	.155	-.20	.136	-.481	.081

groups 1 + 3 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SHOWEMOT	told+affective	15	.73	.458	.118
	not told + affective	15	.73	.458	.118

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
SHOW EMOT	Equal variances assumed	.000	1.000	.000	28	1.000	.00	.167	-.342	.342
	Equal variances not assumed			.000	28.000	1.000	.00	.167	-.342	.342

groups 1 + 4

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SHOWEMOT	told+affective	15	.73	.458	.118
	not told + not affective	15	.47	.516	.133

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
SHOW EMOT	Equal variances assumed	3.646	.067	1.497	28	.146	.27	.178	-.098	.632
	Equal variances not assumed			1.497	27.603	.146	.27	.178	-.099	.632

groups 2 + 3

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SHOWEMOT	told + not affective	15	.93	.258	.067
	not told + affective	15	.73	.458	.118

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
SHOW EMOT	Equal variances assumed	11.146	.002	1.474	28	.152	.20	.136	-.078	.478
	Equal variances not assumed			1.474	22.090	.155	.20	.136	-.081	.481

groups 2 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SHOWEMOT	told + not affective	15	.93	.258	.067
	not told + not affective	15	.47	.516	.133

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SHOWEMOT	Equal variances assumed	40.786	.000	3.130	28	.004	.47	.149	.161	.772
	Equal variances not assumed			3.130	20.588	.005	.47	.149	.156	.777

groups 3 + 4 **Group Statistics**

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
SHOWEMOT	not told + affective	15	.73	.458	.118
	not told + not affective	15	.47	.516	.133

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
SHOWEMOT	Equal variances assumed	3.646	.067	1.497	28	.146	.27	.178	-.098	.632
	Equal variances not assumed			1.497	27.603	.146	.27	.178	-.099	.632

Appendix G. Qualitative Analysis

These qualitative questions were asked after the experiment and a themed analysis was carried out, using colour coded highlighting of text. Summarised results are presented in chapter 5.

During the 'word ladder' game, did you feel any emotions? 49 yes 6 no 5 no response
If so what did you feel?

Overall 30 pos connected terms 72 neg connected. Many related to usability and performance issues 33 reasons given connected with gameplay
Groups told the truth (ie 1 and 4) contained only ones to say they did not feel emotions (8 out of 30)

Group 1 Told and Reactive Yes = 11 No = 2 15 neg connected terms 7 pos connected terms
8 reasons connected with gameplay

x
x
1 Frustration
1 I feel a little nervous and exciting
1 my english is not really good - feel strong when make very silly mistakes
1 puzzlement irritation exuberance
1 determined, slightly irritated
1 I felt nervous before the game and stupid when my mind went blank on the first clue. My confidence grew as I got better at the game
1 confusing some pics didn't quite reflect the word needed
1 frustrated
x
1 feel excited when I get the correct answer and always look at the clock
x
1 annoyed
1 annoyed I got stuck on one of the clues

Group 2 Told and Standard Yes = 14 No = 0 26 neg connected terms 7 pos connected terms
9 reasons connected with gameplay

1 I felt a little disappointed that I could not do the second ladder game
1 Well the computer was very slow and sometime takes time to open a window or positioning the cursor
1 Im so bad at crosswords - slight unease
1 frustration + irritability along with other negative emotions
1 Happy Frustrated
1 thoughtful, knew I should know the answers
1 felt quite tense, thinking very hard, competitative
1 a little frustration when I got stuck - followed by amusement at myself because it is only a game
1 enjoyment when I got it right, frustration & embarrassment when I couldn't think of the right answer
1 confused
1 slightly apprehensive - wasn't sure what was coming next
1 I felt frustrated sometimes whan the mouse didn't work but felt happy when I solved a problem
x
1 annoyed that I couldn't think of the word 'dues' which I didn't actually know
1 stressed a little bit

Group 3 Not Told and Reactive Yes = 15 No = 0 18 neg connected terms 9 pos connected terms 10 reasons connected with gameplay

1 A little bit nervous cos I was not familiar with this game. And interested cos I knew some novel meaning of some words
1 a little bit of anxiety when not able to produce an idea soon
1 I'm not sure I'm just waking up
1 little confusion
1 I felt really excited and interested
1 felt really good when I got the words right and when I understood the game properly
1 bit irritated
1 slightly frustrated and anxious but happy when I got the answer right
1 anger
1 I felt anxious that it would be really hard and beyond me and that I may look stupid
1 excited, fear at screwing up
1 frustration - speed of system
1 anoyed when the picture didn't help and neither did the clue and the help button wouldn't work
1 I felt irritated and annoyed when I couldn't get a word

Group 4 Not Told and Standard Yes = 9 No = 6 13 neg connected terms 7 pos connected terms 6 reasons connected with gameplay

0
0
1 I feel I need to improve my vocabulary a lot and I got nervous as the game went on.
1 I felt happy and excited
1 feeling ws like that I am doing some suspense game but was interesting
0
1 I felt a little challenged at times when I failed to get the right answers - on the other hand felt happy when getting the answer right
1 I felt confident yet nervous to get everything finished on time (in the time limit provided)
1 annoyed because I couldn't get passed the second stage
0
0
0
1 well pissed off it tricked me
1 disappointment and frustration
1 interested, stuck, puzzled

If you felt any emotions, do you think you showed them in any way? 43 yes 17 no
If so how did you show them?

Overall People in 'not told affective' groups gave least responses and those with control set up (not told and std) gave least responses of all.
2 said intentionally didn't show any & 1 gave up trying;
12 mentioned audible ways - 5 laugh / 3 muttering or mumbling /1 swear /1 sigh /1 exclaim /1 repeat
39 mentions of facial expression – 10 general/ 4 eyes or flitting & gaze or looking around/ 8 laugh or happy or smile or grin/6 frown/1 confused / 2 biting lip / 3 touch face or head/ 1 move head / 2 particular parts moved eg. eyebrows / widen eyes / 1 complex state (a look of deep thought)/ 1 sigh
14 mentions of body movement / posture – 4 general / 3 hyper / agitated fidget/ 1 shake legs / 1 move head side to side / 1 fidget / 3 moving on seat eg leaning back and forward, bounced on seat / 1 look of deep thought
13 mentions of hand movements / gestures – 2 general / 3 moving hands or tapping fingers / 2 typing / hit keys harder/ 5 touch face or head/ 1 mouse moves
22 reasons for displays – 6 relating to inner state or feelings /16 to game play or set up

Group 1 Told and Reactive Yes = 11 No = 4 (24) 2 rejected showing any expression; 10 face; 3 voice; 3 body; 2 typing / fingers; 10 reasons given
0 It didn't get anything out of me - I kept a blank face - so it couldn't tell if I was feeling any emotions anyway. I controlled my face so it wouldn't be able to respond to me.
1 Facial expressions
1 Facial expressions Exclamations
1 I think I became alert with wider eyes gazing on screen and more quick move on fingers
1 I expressed them on my face & act like I felt those emotions
1 facial expressions body language
1 frowning and confused look
1 I was aware of the camera + mumbled to myself. I think this was overtly to show how I felt. I smiled to myself when answering correctly.
1 don't know but I don't generally hide them. I don't like feet and that picture made me laugh
1 typing random words
0
1 more hyper and show I'm happy if I get the correct ans
0
0
1 I gave up trying because it was annoying

Group 2 Told and Standard Yes = 14 No = 1 (34) 1 rejected showing any expression; 13 face; 7 voice; 6 body; 3 touch; 5 reasons given
1 mumbling a bit
1 Yes felt them but I never showed up any expression
1 Muttering to myself
1 by shaking my legs and my facial expressions
1 Facial Expression Body movements
1 probably when got an answer after longer time & was easy
1 facial expressions, kept touching face
1 I laughed - tapped my fingers, swore and touched my face
1 laughter, smiling, sitting back / forward, fidgeting, moving my head to from side to side when thinking & probably biting my lip / mouth.
1 laughed at myself for not being able to understand the picture
1 probably socially
1 by frowning and being agitated
0
1 yeah I was repeating the clue and getting quite annoyed with myself
1 frowned

Group 3 Not Told and Reactive Yes = 11 No = 4 0 rejected showing; (24) 11 face; 1 voice; 4 body; 6 hands/ touch/ typing / gesture; 3 reasons given

1 I was nervous so I moved the mouse casually and I frowned
0
1 With my movements and gesturs
1 move on seat, put hand on head
1 I have a funny grin and generally bounced on my seat
0
1 I was looking around to give inspiraation for answer!
1 sometimes with facial expressions
1 eyebrows, touching forehead, laughing to myself

1 My eyes flitted about and I bit my lip
0
1 moving hands around
1 I hit the keys harder in frustration
0
1 perhaps a look of deep thought

Group 4 Not Told and Standard Yes = 7 No = 8 0 rejected showing; 6 face; 1 voice; 1 body; 2 hand / gesture; 4 reasons given

0
0
1 I frown and sigh
1 Maybe I expressed an emotion of frustration when I could not find the right word at the back of my head
1 I showed them in the form of getting happy by myself or sometimes stressed
0
1 I think I did. I believe I start figitting when I'm not comfortable with the answer. I frown a lot I think when challenged.
1 facial expressions and gestures
0
0
0
1 facial expressions during the game
0
1 maybe by putting my hand on my face
0

Did the word ladder game seem to adapt itself to you in any way? 18 yes 42 no
If so how did it adapt?

Overall 10 thought game adapted or gave clues to make easier / 9 thought own effort or knowledge or adaptation / 3 no idea or not sure how / 8 thought basic game design responsible for ease of play
Told groups: 11 yes 19 no 8 comments about game adapting to make easier
Standard groups : gave least responses (4 concepts all from one person)
Told and reacted group : gave most (9 respondents gave 13 concepts)
'Not told' groups: made more remarks related to game design (5 remarks compared to three in told groups)
Told and reacted group: most responses about game adapting (7 compared to three in 3 other groups put together)

Group 1 Told and Reactive Yes = 8 No = 7 7 thought game adapted to make easier / 3 own effort or knowledge / 1 no idea how / 2 game design
0 If it did respond then it gave me the answer. I do not think it did respond. (DELETED If any then I guess it responded to …)
1 After a few times of pressing the help button it started showing pictures next to the word ladder to help me
0
0
1 no idea
1 messages appeared giving me answers
0
1 It seem to get easier when I'd struggled on a few clues but Im not sure if this just shows my getting better at the game
1 it gave me a clue when I needed onnne I asked for help still couldn't get it and when I typed in a wrong answer it gave me a clue
1 give more hints
0
1 It's easy to follow so I can just follow the instruction they gave me
0
0
1 I started getting into it once I got the hang of it

Group 2 Told and Standard Yes = 3 No = 12 1 thought game adapted and made it easier / 3 own effort or knowledge / 1 basic game design
0
1 Just doing the thing again again.
0
0 n/a
1 Gave out the answers
0
1 found it easier to play as time went on - know where things are on screen
0
0
0
0
0
0
0
0

Group 3 Not Told and Reactive Yes = 6 No = 9 1 thought it gave clues or made play easier / 2 own knowledge / 1 no idea / 4 game design features

1 The picture hints are so vivid and easy to accept
1 It is quite a logical game, rational one. It follows a specified path
1 I don't know / I'm just waking up but I could play another game
0
0
0
1 I got used to the game
0
0
0
1 comfortably. Easy to use & understand
0
0
0
1 extra clues popped up when I was having difficulty with the clues

Group 4 Not Told and Standard Yes = 1 (ticked yes but made ambivalent comment) No = 14 1 game design / 1 own adapt / 1 machine adapt /1 not sure

0
0
0
0
0 no
0
0
1 Yes and no - it seemed to become more understandable as I went through the game. I'm not sure if I or the computer adapted more.
0
0
0
0
0
0
0

Do you believe that the computer recognised any of your emotions? If so what emotions did it recognise and how do you think it recognised them?

Overall most yes responses (7) in 'told and reacted' group In all 17 said yes. 43 said no.
'Not told groups gave least responses and std control group least of all
10 suggested aspects of their behaviour recognised by system – all negatively connotation (confused x2, puzzled, difficult, frustration, nervous, idle, struggle, couldn't get it)
11 suggested mechanisms in application that might be responsible for detection (perspex facia on tower had hidden camera/ timer stopped / user pressing help / time taken / pressure on keys & mouse / reaching 4th rung / clicked more than once / reached second quiz /got stuck / time idle /rec by camera)
10 suggested types of response made (help messages or hints / easier questions x 2/ making up words / changing timer x 2 / gave answers / screen shook)

Group 1 Told and Reactive Yes = 7 No = 8 4 deducted something in their behaviour (confused puzzled difficult) 3 mechanism in application (perspex button / timer stopped / pressing button) 4 responses (help messages / easier questions/ making up words / changing timer)

0
1 wrote "n/a" but wanted to know if perspex button was the recording device for his emotions. "Is that the emotion sensor?"
0
1 I noticed that the time calculator stopped whenever I asked for help. That might be a kind of recognition of my emotions
1 I was confused in between so may be that one
1 when I couldn't figure out the answer and felt puzzled help messages appeared
0
1 As above - it seemed to help me regain confidence by giving easier questions
0
0
0
1 yes I feel difficult they will give me hints immediately
0
1 keep pressing help button and making up words
0

Group 2 Told and Standard Yes = 3 No = 12 2 rec something in their behaviour (frustration confusion) 3 mechanisms (time taken / pressure on keys & mouse / 4th rung 1 response

1 by frequently changing the time left
0
0
0 n/a
1 Frustration - by the amount of time I took to respond to the clue - the pressure on keyboard and mouse
0
0
0
0
0
1 It seemed that by the fourth 'rung' down I'd get slightly more confused than previously each time
0
0
0
0

Group 3 Not Told and Reactive Yes = 4 No = 11 2 rec something in their behaviour (nervous/ idle) 4 mechanisms (clicked more than once /second quiz /got stuck /idle) 3 response

1 I got stuck on a new word and was nervous and clicked for help more than once - it would show me more hints to get it solved
1 In the second quiz I was more anxious and it helped me giving me the answers sometimes
1 I'm not sure but as I got the answers when I really had no idea of the word and I didn't when I was guessing it!
0
0
0
0
0
0
0
0
1 gave me more clues when idle
0
0
0

Group 4 Not Told and Standard Yes = 3 No = 12 2 rec something in their behaviour (struggle / couldn't get it) 1 mechanisms (rec by camera) 2 responses

0
0
0
1 I think my emotions were recognised by the camera which was recording my performance
0 no
0
0
1 perhaps in the second game I was struggling and made an easier 'level' appear
0
0
0
0
1 screen started shaking when I couldn't get.
0
0

Do you believe that the computer responded to any of your emotions? If so how did it respond?

Overall 13 yes 47 no
Told and reacted group gave most 'yes's and most written responses
6 responses relating to recognisiton of state of user (troubled, confused, feel stuck)
4 thought it responded with messages or clues
5 thought it supplied answers
4 thought it adapted questions difficulty (one of these thought it made questions harder)
7 mentioned that they thought responses were related to specific point in game / or time measurement

Group 1 Told and Reactive Yes = 6 No = 9 3 recognised state (trouble, confused, feel stuck) 3 thought it responded with messages or clues /3 it supplied answers / 1 adapted questions 5 thought responses were related to point in game / time taken

0 When concentrating on Task 1 it gave an answer when I pressed help. On task 2 it gave an answer much quicker when I pressed help.
1 After a few times of pressing the help button it started showing pictures next to the word ladder to help me
0
0
1 give answers when I was a bit confused
1 It gave me messages
0 was not aware it did but maybe gave me simpler questions
1 as above
0
0
1 the game recognises your trouble with it - the next word is much harder to guess
1 yes when I feel I don't have any clues at the right time the computer give me more hints and I can get the ans straight away
0
0
0

Group 2 Told and Standard Yes = 2 No = 13 **2** recognised state 1supplied answers 1adapted questions 2related to point in game

0
0
0
0 n/a
1 gave out the answers fter a period of time when I did not appear to know the answer
0
1 the questions appeared to get easier after I got stuck on the 'golf' question
0
0
0
0
0
0
0
0

Group 3 Not Told and Reactive Yes = 3 No = 12 1 state 1 supplied answers
0
1
1 In fact I tried to guess the allgorithm of when the computer gives me the answer but I didn't have enough time and I didn't even fought that you are doing such a project
0
1 sometimes they can sense what or how you are feeling
0
0
0
0
0
0
0
0
0
0

Group 4 Not Told and Standard Yes = 2 No = 13 1 responded with messages 2 adapted questions
0
0
0
0
0 no
0
0
1 by changing the difficulty
0
0
0
1 by giving extra clues and making the puzzles easier
0
0
0

In the future all sorts of computer systems and other electronic devices could be designed to recognise your emotions very efficiently and respond to them in different ways, such as by changing the speed of presentation or altering the level of difficulty or altering the type of information presented. What are your views on this? Overall would you like emotional recognition systems to be developed or not? (1 never to 5 strongly want)

Group 1 Told and Reactive

1 Control 1inevitable 3 complexity 9 positive 5 negative 6 caveat 2 uncertainty
Mean = 3.92 **N =** 13 **SD =** .954 **neg connected terms pos connected terms**
3 Technology is designed to make lives easier. If recent generation of electronic devices do this by recognising emotions then I see it as logical progression.If people recognise that adverts are mood targeting, they can still control what they buy.
5 I think it will be extremenly difficult to do this as human emotions can be very complex. People also show their emotions in different ways. It would be interesting.
4 Difficult to answer because it would be desirable in some circumstances but not others
3 I don't like that I don't want the computer to change my life bery much
4 It depends on that presentation because it could be good way but I'm not sure what could be there. Because it can't know in what thing we expressed our emotions it can create more confusions. Should be developed in proper manner that it can't make mistake.
4 Honestly I'm not sure. May be good if used for learning
2 very useful and satisfying
5 This can surely only be a good thing. Confidence is extremely important and if computers can help build this in people by tolorating and adjusting it's a good thing
3 Cool
5 sounds useful
4 I believe we would be very pleased that computer systems and other electronic devices are able to interact with us in a much more interesting sense
5 It would be good idea coz it seemed more flexible and more 'human-being'. I won't think computer game is just a game, I will feel like I'm playing wiv my friend
x I don't av none
x
4 a lot of people will make a lot of money from it

Group 2 Told and Standard

6 Control 0 inevitable 1 complexity 16 positive 9 negative 11 caveat 2 uncertainty
Mean = 4.00 **N =** 15 **SD =** .926 **neg connected terms pos connected terms**
5 It is a good idea to use adaptive computer tests and games. The assessment will be more correct
5 It should not happen in the usual use of a computer because it divert the concentration of the user & feel frustrated
4 Good however in order to do this we must integrate AI essentially allowing a computer extra sensory perception and differential thinking. Decisions should not be based on a computer interface
5 It would be useful especially in educational settings
4 It can only help people therefore it's a good idea and will save time and money hopefully
4 Can all be really useful in right situations. Important to keep researching with introduction of new technologies. Prospects are exciting.
5 I think it would encourage people to play / use devices for longer periods of time
3 I think it is pretty scary that computers are becoming more 'human' and can be adapted to recognise human emotion and error - it fascinates me because I don't understand it!!
3 Perhaps this will be useful for children's games (to aid learning). It would be useful and clever for target advertising to your mood.
2 Would be clever but intrusive way of effecting someones emotions as well
4 Interesting thought which could have several advantages but does throw up some ethical questions about how much 'free choice' a human has.
3 This could be interesting and allow the level of challenge to be set right
4 I regard this to be a very good way forward as it will computers more user friendly for all types of users as it will cater for each user individually
5 It's a really good idea
4

Group 3 Not Told and Reactive

1 Control 2 nevitable 1 complexity 12 positive 3 negative 2 caveat 0 uncertainty
Mean = 4.00 **N =** 15 **SD =** .756 **neg connected terms pos connected terms**
4 It is likely to happen because the tech is improving every day but I don't think it's good because it might have some bad effect on the human imagination
4 Well I think this is a natural step of technology since computers and machines are becoming part of our lives
5 I'm really interested on it and I want to learn more about it. In fact I had some similar ideas but not common with games
4 That is fine - to get feedback from some thoughts
5 That sounds very interesting it would know what kind of mood you are in and change itself for you
4 It's an interesting way to work with computers
4 Slightly good idea. Moreover the computer might be inaccurate about how you really feel + make the user feel annoyed
4 I feel this would be very useful and enable people who do not feel very confident using a computer to gain confidence
3 I feel that the difficulty level should not be changed. We should learn to adapt.
5 I think that used in the right way to a good level then it is a positive thing that could be used it lots of useful ways
5 I think this excellent. Truly man has come along way if this is to be achieved
3 This is good as everyone has a different level of adaptability to tasks / environment
3 Brilliant but they'll never be human
4 I don't see how they could do this
3 indifferent

Group 4 Not Told and Standard

3 Control 1 inevitable 1 complexity 13 positive 1 negative 5 caveat 1 uncertainty
Mean = 3.77 **N = 13** **SD =** 1.092 **neg connected terms pos connected terms**
4 I think it'll be good because the way of information presented influence people's way of thinking a lot. Therefore altering the type of information presented can be efficient.
4 I think it's a brilliant idea and I'm sure it will make the user feel more at ease and comfortable
5 I don't think that computer will ever be possible to recognise human emotions. It is just a dumb terminal and its only human brain which utilise it, but if so, it will be wonderful
5 I think this is good b/c different peoples have different capabilities, and computer has to be intelligent enough to cope with it
2 I guess a more interactive experience like this may prove very valuable especially in the education field where a computer (like a human teacher) would adjust to the level of understanding of a particular individual.
5 I'd be very interested in these devices, music players for instance that adapted to emotions to play your ideal song
5 I think this is fantastic however it should only be done to an extent. For example someone who has a lack of knowledge and is will to learn will not benefit from this if level of difficulty is too easy
3 Technology is increasing rapidly so I am not surprised
3 I believe that it might be interesting although I have no views on it really
4 It will be very helpful especially to people who are unfamiliar with computers
x
3 They should stay how they are
2 Would be very helpful and make the computer programmes easier to use
4 I think this is a good idea and its very interesting
x

	Control	**inevitable**	**complexity**	**positive**	**negative**	**caveat**	**uncertainty**
Group 1	1	1	3	9	5	6	2
Group 2	6	0	1	16	9	11	2
Group 3	1	2	1	12	3	2	0
Group 4	3	1	1	13	1	5	1
Overall	11	4	6	50	18	24	5

Overall would you like emotional recognition systems to be developed or not?
(1 never to 5 strongly want)

Like most new developments, emotional computer systems could greatly improve services, for example: e-learning courses. They could also be used for commercial and other purposes, for example to target advertising to your mood. Overall would you like emotional recognition systems to be developed or not?

1 Feel strongly - they should never be developed	2 Feel mildly - on the whole better if not developed	3 Don't care if developed or not	4 Feel mildly - best if developed	5 Feel strongly - they should be developed

Overall
None said never – only 4 against it, 13 don't care and there were 39 = (nearly two thirds) (vast majority) in favour
Mean 3.93 N56 SD 0.91
2 misled groups slightly more in favour, standard group least –but no significance

Group differences

	1	**2**	**3**	**4**	**5**	**x**
Group 1 Told and Reactive	0	1	3	5	4	2
Group 2 Told and Standard	0	1	3	6	5	0
Group 3 Not Told and Reactive	0	0	4	7	4	0
Group 4 Not Told and Standard	0	2	3	4	4	2
OVERALL	0	4	13	22	17	2

Group 1 told+affective 9 for 1 against Mean 3.92 SD 0.95	Group 2 told + not affective 11 for 1 against Mean 4.0 SD 0.93	Told groups 20 for 2 against	Misled groups 22 for 2 against
Group 3 not told + affective 11 for 1 against Mean 4.0 SD 0.76	Group 4 not told + not affective 8 for 2 against Mean 3.77 SD 1.09	Acted affective groups 20 for 2 against	

Report

should develop emotion systems 1=no to 5= definitely

group	Mean	N	Std. Deviation
told+affective	3.92	13	.954
told + not affective	4.00	15	.926
not told + affective	4.00	15	.756
not told + not affective	3.77	13	1.092
Total	3.93	56	.912

10. References

Abdallah, S., Thompson, S., Michaelson, J., Marks, N. & Steuer N. (2009). The (un)happy planet index 2.0. *NEF thinktank centre for Wellbeing website.* Retrieved 20 September, 2009, from http://www.happyplanetindex.org/.

Achuff, P. (2001). The Lateralization of Emotion. *Brain and Mind Electronic Magazine on Neuroscience*. Retrieved 20 September, 2009, from http://www.cerebromente.org.br/n14/mente/lateralization.htm.

Adolph R., Cahill, L., Schul, R. & Babinsky R. (1997). Impaired declarative memory for emotional material following bilateral amygdala damage in humans. *Learn. Mem.* 4, 291-300.

Alsmeyer, M., Luckin, R. & Good, J. (2008). Developing a novel interface for capturing self reports of affect. In: *Extended Abstracts of CHI 2008,* 2883–2888.

Argyle, M. (1996). *Bodily communication*. London: Routledge.

Arnold, M. B. (1960). *Emotion and personality*. New York: Columbia University Press.

Atlas Ti (2009) *Atlas Ti website.* Retrieved 20 September, 2009, from http://www.atlasti.com/.

Austin, J. L. (1962). *How to do things with words?* 2nd ed. Cambridge, Mass: Harvard University Press.

Averill, J. R. (1980). A constructivist view of emotion. In: Plutchik R. & Kellerman H. (eds.) *Emotion: Theory, research and experience: 1. Theories of emotion.* New York: Academic Press, 305-339. Reprinted in Halberstadt A. & Ellyson, S. L. (eds.) (1990). *Social psychology readings: A century of research.* New York: McGraw-Hill, 143-156.

Bakeman, R. (2000). Behavioral observation and coding. In: Reis H. T. & Judge C. M. (eds.), *Handbook of research methods in social and personality psychology* (pp. 138-159). New York: Cambridge University Press.

Bakeman, R. & Gottman, J.M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge, UK: Cambridge University Press.

Bakeman, R. & Quera V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ*. New York: Cambridge University Press.

Baron-Cohen, S., (2003). *Mind reading: Emotions library user guide version 1.2*. London and New York: Jessica Kingsley Publishers. Retrieved 20 September, 2009, from http://www.jkp.com/mindreading/documents/MindReadingEmotionsLibraryUserGuide.pdf.

Baron Cohen, S. (2009). *The Transporters Website* Retrieved 20 September, 2009, from http://www.thetransporters.com/index.html.

Baron-Cohen, S., Golan, O., Chapman, E. & Granader, Y. (2007). Transported into a world of emotion *The Psychologist,* 20:2, 76-77.

Baron-Cohen, S., Golan, O., Wheelwright, S. & Hill, J. (2004). *Mind reading: The interactive guide to emotions.* London: Jessica Kingsley Publishers.

Baron-Cohen, S., Knickmeyer, R. & Belmonte, M. (2005). Sex differences in the brain: implications for explaining autism. *Science,* 310, 819-823.

Barret, L. F., Robin, L., Pietromonaco, P. & Eyssell, K. (1998). Are women the over emotional sex? ,Evidence from emotional experiences in social context. *Cognition and Emotion*, 12, 555-578.

Bartneck, C. (2002). Integrating the OCC model of emotions in embodied characters. In: *Proceedings of the workshop on virtual conversational characters: Applications, methods, and research challenges.* Retrieved 20 September, 2009, from http://www.vhml.org/workshops/HF2002/papers/bartneck/bartneck.pdf.

Bartneck, C., & Lyons, M. J. (2009). Facial Expression Analysis, Modeling and Synthesis: Overcoming the Limitations of Artificial Intelligence with the Art of the Soluble. In Vallverdu J. & Casacuberta D. (Eds.), Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence IGI Global, 33-53

Baum, L. F. (1900). *The Wonderful Wizard of Oz.* Chicago: George M. Hill Company.

Bavelas, J. B., Black, A., Chovil, N. & Mullett, J. (1990). Truths, lies, and equivocations: The effects of conflicting goals on discourse. *Journal of Language and Social Psychology*, 9, 129-155.

Beck, A.T. (2006). *Depression: causes and treatment*. Philadelphia: University of Pennsylvania Press.

Belavkin, R. V. (2001). The Role of Emotion in Problem Solving. In: *Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing,* York, England: Heslington, 49-57.

Bell, M.A., & Wolfe, C. D. (2004). Emotion and cognition: An intricately bound developmental process. *Child Development*, 75, 366-370.

Benedek, J., & Hazlett, R. (2005). Incorporating facial EMG emotion measures as feedback in the software design process. In: *Proc. Human Computer Interaction Consortium.* Colorado, February 2005.

Bernstein, D. A., Clarke-Stewart, A., Penner, L. A., Roy, E. J., & Wickens, C. D. (2000). *Psychology* 5th ed. Boston, MA: Houghton Mifflin Company.

Bianchi-Berthouze, N. (2008). Using motion capture to recognize affective states in humans. In: *Symposium on Measuring behavior using motion capture, International Conference on Methods and Techniques in Behavioral Research '08,* 28-29.

Bickmore T. & Cassell, J. (2001). Relational agents: a model and implementation of building user trust. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems,* CHI, 31 March - 5 April Seattle, 2001, Washington: ACM, 396-403.

Bickmore, T. Schulman, D. & Sidner C. (2009), Issues in Designing Agents for Long Term Behavior Change CHI 2009 workshop on Engagement by Design. Retrieved 20 September, 2009, from http://www.ccs.neu.edu/home/schulman/

Bird, S. & Harrington, J. (Eds.) *Special issue of Speech Communication on Speech Annotation and Corpus Tools*. 33:1-2. Retrieved 20 September, 2009, from http://www.ldc.upenn.edu/annotation/specom.html.

Black, C., (2008). *Working for a healthier tomorrow.* Retrieved 20 September, 2009, from http://www.workingforhealth.gov.uk/Carol-Blacks-Review/

Blanchard-Fields, F., Stein, R. & Watson, T. L. (2004). Age differences in emotion regulation strategies in handling everyday problems. *Journal of Gerontology: Psychological Sciences,* 59B, 261-269.

Blythe, M., Monk, A., Overbeeke, K. & Wright, P. (eds.) (2003). *Funology: From usability to enjoyment.* Boston: Kluwer Academic Publisher.

Boehner, K., DePaula, R. Dourish P. & Sengers, P. (2005). Affect: from information to interaction. In: Bertelsen, O. W., Bouvin, N. O., Krogh P. G. & Kyng, M. (eds.) *CC'05 Proceedings of the 4th Decennial Conference on Critical Computing: between Sense and Sensibility,* Aarhus, Denmark, 20 – 24 August, 2005 New York, NY: ACM, 59-68.

Boehner, K., DePaula, R., Dourish, P., & Sengers, P. (2007). How emotion is made and measured. *Int. J. Hum.-Comput. Stud.,* 65:4, Apr. 2007, 275-291.

Boehner, K., Sengers, P. & Warner, S. (2008). Interfaces with the ineffable: Meeting aesthetic experience on its own terms. *ACM Trans. Comput.-Hum. Interact,* 15:3, 1-29.

Borod, J. C., Andelman, F., Obler, L. K., Tweedy, J. R. & Welkowitz, J. (1992). Right hemisphere specialization for the appreciation of emotional words and sentences: evidence from stroke patients. *Neuropsychologia* 30, 827 -844.

Bosse, T., Gratch, J., Hoorn, J. F.,Pontier, M. A., & Siddiqui, G. F. (2010). Comparing Three Computational Models of Affect. In: *Proceedings of the 8th International Conference on Practical Applications of Agents and Multi-Agent Systems*, PAAMS'10, 2010. To appear. Preliminary version retrieved 20 September, 2009, from http://www.few.vu.nl/~ghazanfa/papers/PAAMS10.pdf.

Bradley, M. M. & Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of Behav Ther Exp Psychiatry*, 25, 49-59.

Bradley, M. M. & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In: Coan J. A. & Allen J. J. B. (eds.) *Handbook of Emotion Elicitation and Assessment.* New York: Cambridge University Press, 29-46.

Bradley, M. M., Greenwald, M. K. & Petry, M. C. (1992). Remembering pictures: pleasure and arousal in memory. *Journal of Experimental Psychology*, *Learning, Memory, and Cognition*, 18, 379-390.

Breazeal, C. (2000). Sociable Machines: Expressive Social Exchange Between Humans and Robots. dissertation, Department of Electrical Engineering and Computer Science, MIT. Retrieved 20 September, 2009, from http://groups.csail.mit.edu/lbr/hrg/2000/phd.pdf.

Breazeal, C.L. (2002). *Designing sociable robots.* MIT Press.

Briggs, J. L. (1970). *Never in anger: Portrait of an eskimo family.* Cambridge, MA: Harvard University Press.

Brinkman, W. & Fine, N. (2005). Towards customized emotional design: an explorative study of user personality and user interface skin preferences. In: *Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics,* Chania, Greece, 29 September - 01 October, 2005. University of Athens: ACM International Conference Proceeding Series, 132, 107-114.

Brody, L. (1999). *Gender, emotion and the family*. Massachusetts: Harvard University Press.

Broersma, M. (2003). Next-gen PlayStation to sense user emotions. *ZDNetNEWS* Retrieved 20 September, 2009, from http://news.zdnet.co.uk/emergingtech/0,1000000183,39118200,00.htm.

Brown, G., & Yule, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.

Brown, J. & Silzer, P. J. (2001). Miffed, upset, angry, or furious?: translating emotion words. *ATA Chronicle,* 30:7, 40-42.

Brown, P. & Levinson, S. C. (1987). *Politeness: Some universals in language usage.* Cambridge: Cambridge University Press.

Brown, R. & Kulik, J. (1977). Flashbulb memories. *Cognition,* 5, 73–99.

Brundell, P., Tennent, P., Greenhalgh, C., Knight, D., Crabtree, A., O'Malley, C., Ainsworth, S., Clarke, D., Carter R. & Adolphs, S. (2008). Digital Replay System (DRS) - a tool for interaction analysis. In: *Proceedings of the 2008 International Conference on Learning Sciences*, June 23-24, Utrecht, ICSL. Retrieved 20 September, 2009, from http://www.mrl.nott.ac.uk/~axc/DReSS_Outputs/ICLS_2008.pdf .

Bullington, J. (2005). 'Affective' computing and emotion recognition systems: the future of biometric surveillance? In: *InfoSecCD '05. Proceedings of the 2nd Annual Conference on Information Security Curriculum Development.* Kennesaw, Georgia, 23 – 24 September, 2005. New York, NY: ACM, 95-99.

Bunting, M. (2004). *Willing slaves: How the overwork culture is ruining our lives.* London: Harper Collins.

Burman, C. (2002). *Propopagnosia website.* Retrieved 20 September, 2009, from http://www.prosopagnosia.com/.

Business Wire, May 1, (2002). *NetRatings Founder Dave Toth Joins Board At TV Technology Innovator Predictive Networks* Retrieved 20 September, 2009, from http://findarticles.com/p/articles/mi_m0EIN/is_2002_May_1/ai_85366944/.

Cahill, L. & McGaugh, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition*, 4, 410-421.

Cam Studio. (2009). *Cam Studio software website* Retrieved 20 September, 2009, from http://camstudio.org/.

Cañamero, L. & Fredslund, J. (2001). I show you how I like you - can you read it in my face? In: *IEEE Transactions on Systems, Man, and Cybernetics, Part A,* 31:5, 454-459.

Casa, V. D., Hofer, I., Weiner, I. & Feldon, J. (1998). The effects of smoking on acoustic pre-pulse inhibition in healthy men and women. *Psychopharmacology*, 137, 362-368.

Catanzaro, S. J. & Mearns, J. (1990). Measuring generalized expectancies for negative mood regulation: initial scale development and implications. *Journal of Personality Assessment,* 54, 546-563.

Chambers dictionary. (2004). 10th Ed. Harrap Publishers Ltd.

Changing Minds. (2010). *Changing Minds website.* Retreived 10 May 2010 from http://changingminds.org/.

Chapanis, A. (1965). Words, words, words. *Human Factors,* 7, 1-17.

Charles, S. T., Mather, M. & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, 132, 310-324.

Charles, S. T. & Carstensen, L.L. (2007). Emotion regulation and aging. In: Gross, J. J. (ed.) *Handbook of emotion regulation.* New York: Guilford Press, 307-327.

Charlesworth, W. R., & Kreutzer, M. A. (1973). Facial Expressions of Infants and Children. In: Ekman, P. (ed.) *Darwin and Facial Expressions: A Century of Research in Review*. New York: Academic Press, 91-168.

Chen, D. & Haviland-Jones, J. (2000). Human olfactory communication of emotion. *Percept Mot Skills,* Dec. 91,3:1, 771-81.

Chess, S. & Thomas A. (1996). *Temperament.* Routledge.

Christianson, S. A. & Loftus, E. F. (1990). Some characteristics of people's traumatic memories. *Bulletin of the Psychonomic Society*, 28:3, 195-8.

Chwalisz, K., Diener, E. & Gallagher D. (1988). Autonomic arousal feedback and emotional experience: Evidence from the spinal cord injured. *Journal of Personality and Social Psychology*, 54, 820-828.

Clavel, C., Vasilescu, I., Devillers, L. & Richard, G. A. (2008). Fear-type emotions recognition for future audio-based surveillance systems. *Speech Communication,* 50, 487-503.

Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, 53, 751–766.

Cockton, G. & Korhonen, P. (2003). HI 2003 Conference welcome speech. *New Horizon website*. Retrieved 20 September, 2009, from http://www.chi2003.org/theme.html.

Cockton, G. (2004). Doing to be: Multiple routes to affective interaction. *Interacting with Computers,* 16:4, 683-691.

Cockton, G. (2009). When and Why Feelings and Impressions Matter in Interaction Design (Invited Keynote). In: *Kansei 2009: Interfejs Uzytkownika - Kansei w praktyce, CD Rom Proceedings*, Warszawa, Poland. Retrieved 10 May 2010 from http://www.cs.tut.fi/ihte/projects/suxes/pdf/Cockton_Kansei%202009%20Keynote.pdf

Cohn, M. A., Fredrickson, B. L., Brown, S. L., Mikels, J. A. & Conway, A. M. (2009). Happiness unpacked: Positive emotions increase life satisfaction by building resilience. *Emotion,* 9:3, 361-368.

Cole, P. M., Jenkins, P. A. & Shott C. T. (1989). Spontaneous expressive control in blind and sighted children. *Child Development*, 60:3, Jun., 683-688.

Cole, P. M., Martin, S. E. & Dennis, T. A. (2004). Emotion regulation as a scientific construct: methodological challenges and directions for child development research. *Child Dev*, Mar-Apr. 75:2, 317-33.

Companions (2009). *Companions project website*. Retrieved 20 September, 2009, from http://www.companions-project.org/research/relatedprojects.cfm.

Consolvo, S., Klasnja, P., McDonald, D. W., and Landay, J. A. (2009). Goal-setting considerations for persuasive technologies that encourage physical activity. In: *Proceedings of the 4th international Conference on Persuasive Technology,* Claremont, California, 26 – 29 April, 2009, Persuasive '09, Vol. 350, ACM, New York, NY, 1-8.

Cornelius, R. R. (1996). *The science of emotion. Research and tradition in the psychology of emotion.* Upple Saddle River, NJ: Prentice-Hall.

Cornelius, R. R. (2000). Theoretical approaches to emotion. *SpeechEmotion,* 3-10.

Cowie, R. (2007). Humaine Deliverable. *D3i. Emotional life: Terminological and conceptual clarification*s. (2009). Retrieved 20 September, 2009, from http://emotion-research.net/projects/humaine/deliverables/D3i%20final.pdf.

Cowie, R., Douglas-Cowie, E., Savvidou, S., Mcmahon, E., Sawey M. & Schrööder M. (2000). 'Feeltrace': an instrument for recording perceived emotion in real time. In: *Proceedings of the ISCA Workshop on Speech and Emotion,* ISCA, 19-24.

Creed, C. & Beale, R. (2006). Engaging Experiences with Emotional Virtual Therapists. In: *International Design and Engagability Conference @ NordiCHI, 2006,* Oslo: IDEC'06.

Creed, C. P. L. & Beale, R. (2008). Abusive interactions with embodied agents, *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, Special Issue on Abuse and Misuse of Interactive Technologies*, 9:3, 481-503.

Cunningham W.A., Raye, C. L. & Johnson, M. K. (2004). Implicit and explicit evaluation: FMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience,* 16, 1717–29.

D'Mello S. & Graesser, A. (2007). Mind and Body: Dialogue and Posture for Affect Detection in Learning Environments. In: *Proceedings of International Conference on Artificial Intelligence in Education,* Los Angeles. Retrieved 20 September, 2009, from http://www.autotutor.org/publications/newpapers/dmello-aied07.pdf.

Dahlbäck, N. & Jönsson, A. (1989). Empirical studies of discourse representations for natural language interfaces. In: *Proceedings of the Fourth Conference on European Chapter of the Association for Computational Linguistics, European Chapter Meeting of the ACL,* Manchester, England, 10–12 April, 1989, Morristown, NJ: Association For Computational Linguistics, 291-298.

Dahlbäck, N., Jönsson, A. & Ahrenberg, L. (1993). Wizard of oz studies - why and how. *Knowledge-Based Systems,* 6:4. Dec., 258-266.

Damasio, A. R. (1994). *Descartes' error: emotion, reason, and the human brain.* New York: Grosset-Putnam.

Darwin, C. (1872). *The expression of the emotions in man and animals*, London: John Murray. Retrieved 20 September, 2009, from http://darwin-online.org.uk/EditorialIntroductions/Freeman_TheExpressionoftheEmotions.html.

Dautenhahn, K. (1994). Trying to imitate - a step towards releasing robots from social isolation. In: Gaussier P. & Nicoud, J.D. (eds.) *Proceedings From Perception to Action Conference*, Lausanne, Switzerland, 7-9 September, 1994, IEEE Computer Society Press, 290–301.

Davidson, R. J. (1992). Emotion and affective style: hemispheric substrates. *Psyc Sci.* 3, 39-43.

Davidson, R. J. & Fox, N. A. (1982). Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants. *Science,* 218, 1235-1237.

Davis, S. F. & Palladino J. J. (2000). *Psychology* 3rd Ed. Upper Saddle River, NJ: Prentice-Hall, Inc.

De Melo, C. & Gratch, J. (2009). Expression of Emotions using Wrinkles; Blushing; Sweating and Tears. In: *Proc. 9th International Conference on Intelligent Virtual Agents,* Amsterdam. 2009. Retrieved 20 September, 2009, from http://people.ict.usc.edu/~gratch/papers/predict-iva08.pdf

De Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V. & De Carolis, B. (2003). From Greta's Mind to her Face: Modeling the Dynamics of Affective States in a Conversational Embodied Agent. *The International Journal of Human-Computer Studies*, *Special Issue on 'Applications of Affective Computing in Human-Computer Interaction'.* 59:1-2, July 2003, 81-118.

De Wolff, M. S. & van Ljzendoorn, M. H. (1997). Sensitivity and attachment: a meta-analysis on parental antecedents of infant attachment. *Child Dev*, Aug. 68:4, 571-91.

Dillon A. & Jobst, J. (2005). Multimedia learning with hypermedia. In Mayer, R.E. (ed.) *The Cambridge handbook of multimedia learning.* New York: Cambridge University Press, 569-588.

Dimber, U. & Petterson, M. (2000). Facial reactions to happy and angry facial expressions: Evidence for right hemisphere dominance. *Psychophysiology,* 37:5, (The International Journal of the Society for Psychophysiological Research) Cambridge University Press, 693-696.

Dimberg, U. & Lunquist, L. (1990). Gender differences in facial reactions to facial expressions. *Biological Psychology*, 30, 151-159.

Dolcos, F., LaBar, K. S. & Cabeza, R. (2005). Remembering one year later: role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories. In: *Proc Natl. Acad. Sci.,* Feb. 15, 102:7, 2626-2631.

Dyer, C. (2006). *Research in psychology: A practical guide to methods and statistics.* Oxford: Blackwell

Eich, E. & Macaulay, D. (2006). Cognitive and clinical perspectives on mood dependent memory. In: Forgas, J. P. (ed.) *Affect in social thinking and behaviour.* New York: Psychology Press, 105-121.

Eid, M. & Diener E. (2001). Norms for experiencing emotions in different cultures: Inter- and intra-national differences. *Journal of Personality & Social Psychology*, 81, 869 – 885.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, 51, 665-697.

Ekman, P. (1982). *Emotion in the human face.* New York: Cambridge University Press.

Ekman, P. (1997). Should we call it expression or communication? *Innovations in Social Science Research*, 10, 333-344.

Ekman, P. (1999). Basic emotions. In: Dalgleish T. & Power M. (eds.) *Handbook of cognition and emotion.* Sussex, UK: John Wiley & Sons, Ltd., 301-320.

Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life.* Times Books.

Ekman P. & Friesen W. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto.

El Kaliouby R. & Robinson P. (2004) Mind reading machines: automated inference of cognitive mental states from video. In: *Proceedings of The IEEE International Conference on Systems, Man and Cybernetics, 2004,* Vol.1, The Hague, 682 – 688.

El Kaliouby, R. & Robinson, P. (2004). Real-time inference of complex mental states from facial expressions and head gestures. In: *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop Vol 10*, Cvprw'04, 27 June – 02 July, 2004, Washington, DC: IEEE Computer Society, 154.

Elfenbein, H., Marsh, A. & Ambady, N. (2002). Emotional Intelligence and the Recognition of Emotion from Facial Expressions. In: Feldman-Barrett, L. & Salovey, P. (eds.) *The Wisdom in Feeling: Psychological Processes in Emotional Intelligence.* Guilford Press.

Erickson, T. & Laff, M. R. 2001. The design of the 'Babble' timeline: a social proxy for visualizing group activity over time. In: *CHI '01 Extended Abstracts on Human Factors in Computing Systems,* Seattle, Washington, 31 March - 05 April, 2001. CHI '01. ACM, New York, NY, 329-330.

Ethofer, T., Van De Ville, D., Scherer, K. & Vuilleumier P. (2009). Decoding of Emotional Information in Voice-Sensitive Cortices. *Current Biology*, 19:12, 1028-1033.

European Commission Green Paper. (2005). *Confronting demographic change: a new solidarity between the generations.* Retrieved 20 September, 2009, from http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexplus!prod!DocNumber&lg=en&type_doc=COMfinal&an_doc=2005&nu_doc=94.

Fagerberg, P., Ståhl, A., & Höök, K. (2004). eMoto: emotionally engaging interaction. *Personal Ubiquitous Comput.,* 8:5. Sep., 377-381.

Fajardo, N. & Moere, A. V. (2008). ExternalEyes: evaluating the visual abstraction of human emotion on a public wearable display device. In: *Proceedings of the 20th Australasian Conference on Computer-Human interaction: Designing For Habitus and Habitat.* Cairns, Australia, December 08 - 12, 2008. OZCHI '08, 287 New York, NY: ACM., 247-250.

Fasel B. & Luettin J. (2003). Automatic facial expression analysis: A survey. *Pattern Recognition*, 36:2, 259-275.

Fehr, B. & Russell, A. J. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, 113, 464-486.

Fleck, R, & Fitzpatrick, G. (2006). Supporting Collaborative Reflection with a Passive Image Capture Device. In: *SupplementaryProc COOP2006*, Carry-le-Rouet, France, 41-48. Retrieved 20 September, 2009, from http://www.sussex.ac.uk/Users/rmmf20/RFCOOPShort.pdf.

Fletcher, L. & Zelionsky, A. (2007). Driver state monitoring to mitigate distraction. In: Faulks, I. J., Regan, M., Stevenson, M., Brown, J., Porter A. & Irwin, J. D. (eds.) *Distracted driving.* Sydney, NSW: Australasian College of Road Safety, 487-523.

Fogg, B. J. (2002). *Persuasive technology: Using computers to change what we think and do.* Morgan Kaufmann.

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. (2007). The world of emotion is not two-dimensional. *Psychological Science,* 18:2, 1050–1057.

Fox, N.A. (1991). If it's not left, it's right. *Am Psych,* 46, 863-872.

Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology*, 2, 300-319.

Fredrickson, B. L. (2000). Cultivating Positive Emotions to Optimize Health and Wellbeing. *Prevention and Treatment 3.* Retrieved 20 September, 2009, from http://www.unc.edu/peplab/publications/cultivating.pdf.

Fredrickson, B. L., Tugade, M. M., Waugh, C. E., & Larkin, G. (2003). What good are positive emotions in crises?: A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *Journal of Personality and Social Psychology*, 84, 365-376.

Freed, K. (2001). OpenTV Signs Predictive for Targeted iTV Advertising in Interactive TV Trade Reports, *Multichannel News International*. Retrieved 20 September, 2009, from http://www.media-visions.com/itv-predictive.html.

Freeman, D.E. (2003). *Creating emotion in games: the art and craft of emotioneering*. New Riders.

Fridja, N. H. (1986). *The emotions.* Cambridge University Press.

Fridlund, A. J. (1991). The sociality of solitary smiles: Effects of an implicit audience. *Journal of. Personality and Social Psychology*, 60, 229-240.

Gardner, M. (1996). Word Ladders: Lewis Carroll's Doublets. *The Mathematical Gazette*, 80:487, Centenary Issue, Mar., 1996, 195-198. The Mathematical Association. Retrieved 20 September, 2009, from http://www.jstor.org/stable/3620349

Gelin, P. (2005). Humaine Deliverable. *D4d. Proposals for exemplars and work towards them: signs and signals* Retrieved 20 September, 2009, from http://emotion-research.net/projects/humaine/deliverables/D4d%20final.pdf.

Gerrards-Hesse, A., Spies, K., & Hesse, F.W. (1994). Experimental inductions of emotional states and their effectiveness - A review. *British Journal of Psychology*, 85, 55-78.

Geuens M. & de Pelsmacker P. (2002) Developing a short Affect Intensity Scale, *Psychol Rep.,* 91, 657–670.

Gibberman, S. R. (1991). *Star Trek: An annotated guide to resources on the development, the phenomenon, the people, the television series, the films, the novels, and the recordings.* Jefferson, North Carolina: McFarland.

Gillespie, R. (1991). *Manufacturing knowledge: a history of the Hawthorne experiments.* Cambridge: Cambridge University Press

Gimson, A.C. Cruttenden, A. (2001). *Gimson's Pronunciation of English*, 6th Ed. Arnold Publishers.

Givens, D. B. (2009). *Nonverbal dictionary of gestures, signs and body language cues.* Spokane, Washington: Center for Nonverbal Studies Press, 2004. Retrieved 20 September, 2009, from http://center-for-nonverbal-studies.org/6101.html.

Glaser, B. G. (1998). *Doing grounded theory - Issues and discussions*. Sociology Press.

Goeleven, E., De Raedt, R., Leyman, L. & Verschuere, B. (2008). A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102-141.

Goldberg, L. R. (1992). The development of markers for the Big-five factor structure. *Journal of Personality and Social Psychology*, 59:6, 1216-1229.

Goldman, A. I. (2005). Imitation, mind reading, and simulation. In: Hurley S. & Charter, N. (eds.) *Perspectives on imitation 2.* Cambridge, MA: MIT Press, 79–93.

Goleman, D. (1995). *Emotional intelligence.* New York: Bantam Books.

Goodwin, C. (2004). A competent speaker who can't speak: The social life of aphasia. *Journal of Linguistic Anthropology,* 14:2, 151-170.

Goos, L. & Silverman, I. (2002). Sex Related Factors in the Perception of Threatening Facial Expressions *Journal of nonverbal Behavior*, 26:1, Spring 2002, 27-41.

Gould, J. D. & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM*, 28:3, March 1985, 300-311.

Grand, S. (2004). *Growing up with Lucy: How to build an android in twenty easy steps.* Weidenfeld & Nicolson.

Green, P. & Wei-Haas, L. (1985). The rapid development of user interfaces: Experience with the Wizard of OZ method. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 29:5, 470-475.

Grossman, M. & Wood, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology*, 65, 1010-1022.

Grüter T., Grüter M. & Carbon, C. C. (2008). Neural and genetic foundations of face recognition and prosopagnosia. *J. Neuropsychol.* 2:1, 79–97.

Gunes, H. & Piccardi, M. (2006). A Bimodal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior, In: *Proc. of ICPR 2006 the 18th IAPR International Conference on Pattern Recognition*, *Vol. 1,* 22-24, August 2006, Hong Kong: IEEE Computer Society Press, 1148-1153. 2

Halko, S. & J. A. Kientz. (2009). Personality and Persuasive Technology: An Exploratory Study on Health-Promoting Mobile Applications. In: *Proceedings of Persuasive 2009*, Copenhagen, Denmark. 2010, 150-161.

Hall, J. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. The Johns Hopkins University Press.

Hall, L., Woods, S., Hall, M., & Wolke, D. (2007). Children's Emotional Interpretation of Synthetic Character Interactions. In: *Proceedings of the 2nd international Conference on Affective Computing and intelligent interaction,* Lisbon, Portugal, 12 – 14 September, 2007, Paiva A. C., Prada, R. & Picard, R. W. (eds.) Lecture Notes In Computer Science, Vol. 4738, Springer-Verlag, Berlin, Heidelberg, 642-653.

Harland, C. J., Clark, T. D. & Prance, R. (2002). Electric potential probes - new directions in the remote sensing of the human body *Meas. Sci. Tech.* 13, 163-169.

Hayes, A. (2005) Website on SPSS macros. Retrieved 20 September, 2009, from http://www.comm.ohio-state.edu/ahayes/

Hayes-Roth, B., Ball, G., Lisetti, C., Picard, R. & Stern, A. (1998) Panel on affect and emotion in the user interface. In: *Proc Int. Conf. Intelligent User Interfaces*, 1998, 91-94.

Hazlett, R. (2003). Measurement of user frustration: a biologic approach. In: *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, Ft. Lauderdale, Florida, USA, April 05 - 10, 2003. CHI '03. ACM, New York, NY, 734-735.

Hedlund, J. L. & Viewig, B. W. (1979). The Hamilton rating scale for depression: a comprehensive review. *Journal of Operational Psychiatry,* 10, 149-165.

Hellige, J. B. (1993). *Hemispheric asymmetry.* Cambridge, MA: Harvard University Press.

Hiatt, L. R. (1978). Classification of the emotions. In: Hiatt L. R. (ed.) *Australian aboriginal concepts.* Princeton, NJ: Humanities Press, 182-187.

Hinde, R. A. (1985). Was "The Expression of Emotions" a Misleading Phrase? *Animal Behaviour,* 33, 985–92.

Hittelman, J. H. & Dickes, R. (1979). Sex Differences in Neonatal Eye Contact Time. *Merrill-Palmer Quarterly*, *Jul,* 25:3, 71-84.

Hochschild, A. R. (1983). *The managed heart: Commercialization of human feeling.* University of California Press.

Hofmann, S. G., Suvak, M. & Litz, B. T. (2006). Sex differences in face recognition and influence of facial affect. *Personality and Individual Differences,* June, 40:8, 1683-1690.

Hohmann, G. W. (1966). Some effects of spinal cord lesions on experienced emotional feelings. *Psychophysiology,* 1966: 3, 143-56.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.,* 7:2, Jun. 2000, 174-196.

Hoover Institute. (2009). *Hoover Institute, Stanford University website*, Future Shock: High Technology and the Human Prospect. Peter Robinson 'Uncommon Knowledge' interview with Bill Joy and Ray Kurzweil. Retrieved 20 September, 2009, from http://www.hoover.org/multimedia/uk/3004471.html.

Hornik, R., Risenhoover, N. & Gunnar, M. (1987). The effects of maternal positive, neutral and negative affective communications on infant responses to new toys. *Child Development,* 58, 937–944.

Howard, A. (2006). Positive and negative emotional attractors and intentional change *The Journal of Management Development,* 25:7, 657-670.

Hsee, C. K., Hatfield, E., Carlson, J. & Chemtob, C. (1990). The effect of power on susceptibility to emotional contagion. *Cognition and Emotion,* 4, 327-340.

Hu, B., Moore, P. & Wan, J. (2008). Ontology based mobile monitoring and treatment against depression. *Wireless Communications and Mobile Computing, Special Issue on Pervasive Computing Technology and its Applications*, 1-16.

Humaine toolbox. *(2009). Humaine website: toolbox.* Retrieved 20 September, 2009, from http://emotion-research.net/toolbox/toolbox.

Humaine. (2009). *Humaine website*. Retrieved 20 September, from http://emotion-research.net/projects/humaine/aboutHUMAINE/start-here.

Hurley, S. & Chater, N. (eds.) (2005). *Perspectives on imitation: From neuroscience to social science.* MIT Press.

Hutchby, I. & Wooffitt, R. (1988). *Conversation Analysis*. Polity Press.

Hutchinson-Guest, A. (1989). *Choreographics: a comparison of dance notation systems from the fifteenth century to the present*. Routledge.

Intini, J. (2005). Robo-sapiens rising: Sony, Honda and others are spending millions to put a robot in your house. *Macleans.CA website.* Retrieved 20 September, 2009, from http://www.macleans.ca/article.jsp?content=20050718_109126_109126.

Isbister, K., Höök, K., Laaksolahti, J., and Sharp, M. (2007). The sensual evaluation instrument: Developing a trans-cultural self-report measure of affect. *Int. J. Hum.-Comput. Stud.,* 65:4, April, 2007, 315-328.

Isen, A. M. (2000). Positive affect and decision making. In: Lewis M. & Haviland-Jones, J. (eds.) *Handbook of emotions.* 2nd Ed. New York: Guilford, 417-435.

Isen, A.M., Shalker, T.E., Clark, M. & Karp, L. (1978). Affect, accessibility of material in memory, and behavior: a cognitive loop? *Journal of Personality and Social Psychology.*36, 1–11.

Izard, C. E. (1977). *Human emotions.* New York: Plenum.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review.* 99:3, 561-565.

Izard, C., Kagan, J. & Zajonc, R. (eds.) (1984). *Emotions, cognition and behavior.* New York, NY: Cambridge University Press.

James, W. (1884). What is an emotion? *Mind,* 19, 188-205.

James, W. (1890). *The principles of psychology.* New York: Henry Holt. (Reprinted Bristol: Thoemmes Press, 1999).

Jennings, P.D., McGinnis, D., Lovejoy, S. & Stirling, J. (2000). Valence and arousal ratings for Velten mood induction procedures. *Motiv Emot.,* 24, 285–297.

Jones, M. B. & Jones, D. R. (1995). Preferred pathways of behavioural contagion. *Journal of Psychiatric Research*, 29, 193-209.3

Kaiser, S. & Wehrle, T. (1992). Automated coding of facial behavior in human-computer interactions with FACS, *Journal of Nonverbal Behavior,* 16:2, 67-83.

Kaiser, S. & Wehrle, T. (1996). Situated emotional problem solving in interactive computer games. In: *Proceedings of the VIXth Conference of the International Society for Research on Emotions,* ISRE Publications, 276-280.

Kaiser, S., Wehrle, T. & Schmidt, S. (1998). Emotional episodes, facial expressions, and reported feelings in human-computer interactions. In: Fischer A. H. (ed.) *Proceedings of the Xth Conference of the International Society for Research on Emotions.* Würzburg: ISRE Publications, 82-86.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Trans. Inf. Syst.,* 2:1, Jan. 1984, 26-41.

Kendon, A. & Ferber, A. (1973). A description of some human greetings. In: Michael R. & Crook J. (eds.) *Comparative ecology and behaviour of primates.* London: Academic Press, 591-668.

Kenny, P., Parsons, T., Gratch, J., & Rizzo, A. (2008). Virtual humans for assisted health care. In: Makedon, F. Baillie, L. Pantziou, G. & Maglogiannis, I. (eds.) *PETRA '08, Proceedings of the 1st international Conference on Pervasive Technologies Related To Assistive Environments,* Athens, Greece, July 16-18, 2008. 282. New York, NY: ACM, 1-4.

Kent, N. (2009). *Computer rage website.* Retrieved 20 September, 2009, from http://lap.umd.edu/computer_rage/.

Kiefer, C., Collins, N & Fitzpatrick, G. (2009). Phalanger: Controlling music software with hand movement using a computer vision and machine learning approach. In: *New Interfaces for Musical Expression*, 4- 6 June 2009, Pittsburgh, USA.

Kinect. (2010) *Kinect website.* Retrieved 20 June, 2010, from http://www.xbox.com/en-us/kinect.

Kipp, M. (2001). Anvil - a generic annotation tool for multimodal dialogue. In: *Proceedings of the 7th European Conference on Speech Communication and Technology,* Eurospeech, Aalborg, Denmark, 1367-1370.

Kirsch, D. (1997). *The Sentic Mouse: A tool for measuring emotional valence*, MIT, May 1997. Retrieved 20 September, 2009, from http://www.media.mit.edu/affect/AC_research/projects/sentic_mouse.html.

Kirsch D. (1999). The Affective Tigger: a study on the construction of an emotionally reactive toy. Master's Thesis, MIT, 1999. Retrieved 20 September, 2009, from http://vismod.media.mit.edu//tech-reports/TR-496.pdf

Klein, J., Moon, Y. & Picard, R. W. (2002). This Computer Responds to User Frustration. *Interacting with Computers,* 14:2, 119-140.

Kleinginna, P.R. & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion,* December 5:4, 345-379.

Kleinsmith, A. & Bianchi-Berthouze N. (2007). Recognizing affective dimensions from body posture. *In: Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction, 2007,* Lisbon: ACII'07, 48-58.

Kleinsmith, A., De Silva, P. R. & Bianchi-Berthouze, N. (2006). Cross-cultural differences in recognizing affect from body posture. *Interact. Comput.* 18:6, Dec., 1371-1389.

Knoppel, F. L., Tigelaar, A. S., Bos, D. O., Alofs, T., & Ruttkay, Z. (2008). Trackside DEIRA: a dynamic engaging intelligent reporter agent. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems – Vol 1,* Estoril, Portugal, 12 – 16 May, 2008, International Conference on Autonomous Agents. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 112-119.

Knudsen, M. W., Martin, J. C., Dybkjær, L., Berman, S., Bernsen, N. O., Choukri, K., Heid. U., Kita, S., Mapelli, V., Pelachaud, C., Poggi, I., van Elswijk, G. & Wittenburg, P. (2002). ISLE Natural Interactivity and Multimodality Working Group Deliverable. *D8.1 Survey of NIMM Data resources, Current and Future User Profiles, Markets and User Needs for NIMM Resources.* ISLE website. Retrieved 20 September, 2009, from http://isle.nis.sdu.dk/reports/wp8/.

Koay, K. L. Syrdal, D. S., Walters, M. L. & Dautenhahn K. (2007). Living with Robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In: *Proceedings of the 16th IEEE International Workshop on Robot and Human Interactive Communication.* RO-MAN 2007, South Korea, 564 569.

Koonce, R. (1996). Emotional IQ, a new secret of success? *Training & Development*, 50:2, 19.

Krause, R., Steimer-Krause E. & Ulrich, B. (1992). Use of affect research in dynamic psychotherapy. In: Leuzinger-Bohlebe M., Schneider R, H. & Pfeifer, R. (eds.) *Two butterflies on my head. Psychoanalysis in the interdisciplinary dialogue.* Berlin: Springer, 277-291.

Kraut, R.E. & Johnson, R.E. (1979). Social and emotional messages of smiling: An ethological approach. *J Pers Soc Psychol.* 37, 1539–1553.

Kurzweil, R., (1999). *The age of spiritual machines*. Viking Adult. Retrieved 20 September, 2009, from http://us.penguingroup.com/static/packages/us/kurzweil/excerpts/exmain.htm.

Lal, S.K.L. & Craig, A. (2002). Driver fatigue: Electroencephalography and psychological assessment *Psychophysiology* 39:3, Cambridge Uni Press, 313-321.

Lamb, S. M. (1998). *Pathways of the Brain: The Neurocognitive Basis of Language*. Amsterdam: John Benjamins. Language and the Brain Neurocognitive Linguistics website Retrieved 20 September, 2009, from http://www.ruf.rice.edu/~lngbrain/

Lamb, S. M. (2004). *Language and Reality* Webster, J. (ed) London: Continuum Books.

Lang, P. J., Öhman, A. & Vaitl, D. (1988). *The international affective picture system (photographic slides).* Available from: Center for research in psychophysiology, University of Florida NIMH Media Core website. Retrieved 20 September, 2009, from http://csea.phhp.ufl.edu/Media.html.

Lavigne, J. J., Savoy, S., Clevenger, M. B., Richie, J. E., McDoniel, B., Yoo, S. J., Anslyn, E. V., McDevitt, J. T., Shear, J. B. & Neikirk, D. (1998). Solution-based analysis of multiple analyses by a sensor array: toward the development of an 'electronic tongue. *Journal of the American Chemical Society*, 120, July, 6429-6430.

Lazarus, R. (1991). *Emotion and adaptation.* Oxford University Press.

Lazarus, R. S. (1993). From psychological stress to the emotions: A history of changing outlooks. *Annual Review of Psychology*. 44, 1-21.

Lazarus, R. S. & Folkman, S. (1984) *Stress, appraisal, and coping*. New York: Springer Pub. Co.

Lazarus, R. S., Speisman, J. C., Mordkoff, J. & Davison, L. (1962). A laboratory study of psychological stress produced by a motion picture film. *Psychol Monogr.* 76:34.

Leahu, L., Schwenk, S. & Sengers, P. (2008). Subjective objectivity: negotiating emotional meaning. In: *Proceedings of the 7th ACM Conference on Designing interactive Systems*, DIS '08, Cape Town, South Africa, 25 – 27 February, 2008, New York, NY: ACM, 425-434.

Lear, R. & Palmer, C. (2008). Obesity and the exercise opportunity, a critical dialogue in Physical Education. *Journal of Qualitative Research in Sports Studies*, 2:1, 83-98.

Ledoux, J. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon & Schuster.

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology* 23, 727–38.

Lee, C., Lesh, N., Sidner, C. L., Morency, L., Kapoor, A., and Darrell, T. 2004. Nodding in conversations with a robot. In: *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, Vienna, Austria, 24 - 29April, 2004, CHI '04, ACM, New York, NY, 785-786.

Leeb, R.T. & Rejskind, F.G. (2004). Here's looking at you, kid: A longitudinal study of perceived gender differences in mutual gaze behavior in young infants. *Sex Roles* 50:1/2, 1-14.

Lehto, L., Laaksonen, L., Vilkman, E. & Alku A. (2008). Changes in objective acoustic measurements and subjective voice complaints in call center customer-service advisors during one working day. *Journal of Voice*, 22:2, 164-177.

Levenson, R. W., Ekman, P. & Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology,* Jul, 27:4, 363-84.

Levine, S.P. & Feldman, R.S. (1997). Self presentational goals, self-monitoring, and nonverbal behavior. *Basic and Applied Social Psychology*, 19:4, 505-518.

Lewins, A. & C. Silver. (2007). *Using qualitative software: A step by step guide*. London: Sage Publications.

Lindeman, R. W. & Noma, H. (2007). A classification scheme for multi-sensory augmented reality. In: Spencer, S. N. (ed.) *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology,* VRST '07 Newport Beach, California, 05 – 07 November, 2007, New York, NY: ACM, 175-178.

Lindgaard, G. (2004). Adventurers versus nit-pickers on affective computing. *Interacting with Computers* 16:4, 723-728.2

Lindström, M., Ståhl, A., Höök, K., Sundström, P., Laaksolahti, J., Combetto, M., Taylor, A. & Bresin, R. (2006). Affective Diary - designing for bodily expressiveness and self-reflection. In: *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, Montréal, Québec, Canada, 22 – 27 April, 2006, CHI '06. ACM, New York, NY, 1037-1042.

Lombard, M., Snyder-Duch, J. & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research,* 28, 587-604.

Lombard, M., Snyder-Duch, J. & Bracken, C. C. (2005) Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects. Retrieved 20 September, 2009, from http://www.temple.edu/mmc/reliability/

Longfellow, H. W. (1858). The courtship of Miles Standish VI Priscilla. Retrieved 20 September, 2009, from http://www.portitude.org/literature/longfellow/pt-courtship_of_miles_standish.php.

Luria, A. R. (1973). *The Working Brain.* Basic Books.

Lyons, M., Campbell, R., Plante, A., Coleman, M., Kamachi, M. & Akamatsu, S. (2000). The Noh mask effect: Vertical viewpoint dependence of facial expression perception. In: *Proceedings of the Royal Society of London* B:267, 2239-2245. Retrieved 20 September, 2009, from http://www.kasrl.org/noh_mask.html.

MacGeorge E. L., Gillihan, S. J., Samter, W. & Clark, R. A. (2003). Skill deficit or differential motivation? Testing alternative explanations for gender differences in the provision of emotional support. *Communication Research*, 30:3, 272-303.

Mahlke, S. & Minge, M. (2008). Consideration of Multiple Components of Emotions in Human-Technology Interaction. In: Peter C. & Beale R. (eds.) *Affect and emotion in human-computer interaction: From theory to applications*. Lecture notes in computer science 4868 Berlin, Heidelberg: Springer-Verlag, 51-62.

Mandler, G. (1975). *Mind and emotion*. New York: Wiley.

Mandryk**,** R. L. & Atkins M. S. (2007). A Fuzzy Physiological Approach for Continuously Modeling Emotion During Interaction with Play Environments. *International Journal of Human-Computer Studies*, 6:4, 329-347.

Mandryk, R., Inkpen, K. & Calvert T. (2006). Using Psychophysiological Techniques to Measure User Experience with Entertainment Technologies. *Behaviour and Information Technology, Special Issue on User Experience*, March-April 25:2, 141-58.

Marakas, G. M., Johnson, R. D., & Palmer, J. (2000). A theoretical model of differential social attributions toward computing technology: When the metaphor becomes the model. *International Journal of Human-Computer Studies*, 52:4, 719-750.

Martin, J. R. & White, P. R. R. (2005). *The Appraisal Website*. Retrieved 20 September, 2009, from http://www.grammatics.com/Appraisal/.

Martin, J. R. & White, P. R. R. (2005). *The language of evaluation, appraisal in English*. London & New York: Palgrave Macmillan.

Martin, M. (1990). On the induction of mood. *Clin Psychol Rev*., 10, 669–697.

Mathews, T. J. (1996). A case for increasing negative affect in foreign language classes. *Language Learning Journal*, 13:1, 1753-2167.

Matsumoso, D. & Ekman, P. (2008). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion*, 13, 143-157.

Matsumoto, D. & Kupperbusch, C. (2001). Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience. *Asian Journal of Social Psychology,* 4, 113-131.

Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., Schwalb, J. M. & Kennedy, S. H. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron.* 45:5, March, 651–60.

McFedries, P. (2009). *Wordspy: The word lovers guide to new words*. Retrieved 20 September, 2009, from http://www.wordspy.com/.

McGilloway, S., Cowie, R., Douglas-Cowie, E., Gielen, S., Westerdijk, M. & Stroeve, S. (2000). Automatic recognition of emotion from voice: A rough benchmark. In Cowie, R. Douglas-Cowie, E. & Schroeder M. (eds.), *Speech and emotion: Proceedings of the ISCA workshop,* Newcastle, Co. Down, September 2000, 207-212.

McGinley, M. (2008). Mouse Trax Mouse tracking shareware *Softpedia website*. Retrieved 20 September, 2009, from http://mac.softpedia.com/get/Dashboard-Widgets/Games-Widgets/MouseTrax-widget.shtml.

McGivern, R. F., Andersen, J., Byrd, D., Mutter, K. L. & Reilly, J. (2002). Cognitive efficiency on a match to sample task decreases at the onset of puberty in children *Brain & Cognition*, 50, 73–89.

Mead, M. (1928). Selection from Coming of Age in Samoa. Retrieved 20 September, 2009, from http://faculty.salisbury.edu/~mllewis/Utopia/MeadCoAiS.pdf .

Mehrabian, A. & Ferris, S. R. (1967). Inference of Attitudes from Nonverbal Communication in Two Channels. *Journal of Consulting Psychology* 31:3, 248–252.

Mehrabian, A. & Wiener, M. (1967). Decoding of Inconsistent Communications. *Journal of Personality and Social Psychology* 6:1, 109–114.

Messinger, D. (2009). *Continuous Measurement System website.* Retrieved 20 September, 2009, from http://measurement.psy.miami.edu/cms.phtml.

Minsky, M. (2006). *The Emotion Machine*. (draft version) Retrieved 20 September, 2009, from http://web.media.mit.edu/~minsky/eb5.html.

Mizerski, R. W. & White, J. D. (1986). Understanding and using emotions in advertising. *Journal of Consumer Marketing,* 3:4, 57-69.

Moore, D. J. & Harris, W. D. (1996). Affect intensity and the consumer response to high impact emotional advertising appeals. *Journal of Advertising*, 25:1 Spring, 37-50.

Morency, L. P., de Kok, I. & Gratch. J. (2008). Context-based Recognition during Human Interactions: Automatic Feature Selection and Encoding Dictionary. In: *Proc. 10th International Conference on Multimodal Interfaces* (ICMI 2008), October 2008. Retrieved 20 September, 2009, from

http://people.ict.usc.edu/~gratch/papers/morency-ICMI2008.pdf.

Morency, L. P., de Kok, I. & Gratch. J. (2008). Predicting Listener Backchannels: A Probabilistic Multimodal Approach. *8th International Conference on Intelligent Virtual Agents,* Tokyo, Japan, September 2008. Retrieved 20 September, 2009, from http://people.ict.usc.edu/~gratch/papers/predict-iva08.pdf.

Mori, M. (1970). *Bukimi no tani. The uncanny valley* (Originally in Japanese. MacDorman K. F. & Minato, T. Trans.) Energy, 7:4, 33–35. Retrieved 20 September, 2009, from http://www.androidscience.com/theuncannyvalley/proceedings2005/uncannyvalley.html

Mori, M. (2005). On the Uncanny Valley. In: *Proceedings of the Humanoids-2005 workshop: Views of the Uncanny Valley*. 5 December 2005, Tsukuba, Japan.

Moridis, C. N. &, Economides, A. A. (2008). Toward Computer-aided affective learning systems: A literature review. *Journal of Educational Computing Research,* 39:4. 313-337.

Müller, R. A., Kleinhans, N., Kemmotsu, N., Pierce, K. & Courchesne, E. (2003). Abnormal variability and distribution of functional maps in autism: An fMRI study of visuomotor learning. *Am J Psychiatry,* October, 160, 1847-1862.

Mumme, D. & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a televised scenario. *Child Development*, 74, 221-237.

Nadel, J. (2004). Early imitation and the emergence of a sense of agency. In: Berthouze, L., Kozima, H., Prince, C.G., Sandini, G., Stojanov, G., Metta, G. & Balkenius, C. (eds.) In: *Proceedings Fourth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems,* 117, Genoa, Italy, 15-16.

Nadel, J. & Butterworth, G. (eds.) (1999). *Imitation in infancy. Cambridge studies in cognitive and perceptual development.* New York: Cambridge University Press.

Newton J. (2007). Wellbeing Research: Synthesis Report. *DEFRA website.* Retrieved 20 September, 2009, from http://www.defra.gov.uk/sustainable/government/documents/Wellbeingresearchsynthesisreport.pdf.

Nielsen, J. (1994). Heuristic evaluation. In: Nielsen, J. & Mack, R.L. (eds.) *Usability Inspection Methods*, New York, NY: John Wiley & Sons.

Nielsen, J. (2009) Ten Usability Heuristics. Retrieved 20 September, 2009, from http://www.useit.com/papers/heuristic/heuristic_list.html

Nold, C. (2009). Emotional Cartography, *Bio Mapping website*. Retrieved 20 September, 2009, from http://biomapping.net.

Noldus. (2009). *Face reader website.* Retrieved 20 September, 2009, from http://www.noldus.com/human-behavior-research/products/facereader

Noldus Technology. (2009). The ObserverR, *Noldus Information Technology website*. Retrieved 20 September, 2009, from http://www.noldus.com/products/index.html?observer/index .

Norman, D. A. (2004). *Emotional design: Attractive things work better.* New York: Basic Books.

Norman, M. A. & Thomas, P. J. (1991). Informing HCI design through conversation analysis. *Int. J. Man-Mach. Stud.* 35:2, Sep. 235-250.

Nunn, J. A., Gregory L. J. & Brammer, M. (2002). Functional magnetic resonance imaging of synesthesia: activation of V4/V8 by spoken words. *Nature Neuroscience* 5, 371-375.

NVivo (2009). *NVivo website*. Retrieved 20 September, 2009, from http://www.qsrinternational.com/products_nvivo.aspx.

Nygaard L. C., (2005). Perceptual Integration of Linguistic and nonlinguistic properties of speech. In: D.B. Pisoni & R.E. Remez, (Eds.) *The handbook of speech perception*. Oxford: Blackwell, 390-413.

Oatley, K. (2004). *Emotions: A brief history*. Wiley Blackwell.

Oatley, K. & Johnson-Laird, P.N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion* 1:1, 29–50.

Ochs, E., Graesch, A., Mittman, A., Bradbury T. & Repetti, R. (2006). Video ethnography and ethnoarcheological tracking. In: Pitt-Catsouphes M. & Sweet, S. (eds.) *Handbook of work and family* Mahwah, NJ: Lawrence Erlbaum Associates, 387-410.

Ochsner, K. N., Bunge, S. A., Gross, J. J. & Gabrieli, J.D. (2002). Rethinking feelings: an fMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience,* 14, 1215–29.

Ofcom (2007). *Technology Trends Workshop Report*, December 2007. Retrieved 20 September, 2009, from http://www.ofcom.org.uk/research/technology/events/techtrends07/report07.pdf.

Ofcom, (2009). *The Communications Market 200,* August. Retrieved 20 September, 2009, from http://www.ofcom.org.uk/research/cm/cmr09/.

Ortony, A. (2002). On making believable emotional agents believable. In: R. Trappl, P. Petta & S. Payr, (eds.) *Emotions in humans and artifacts,* Cambridge, MA: MIT Press, 189-212.

Ortony, A., Clore, G. & Collins, A. (1988). *The cognitive structure of emotions.* Cambridge: Cambridge University Press.

Ortony, A. & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review*, 97, 315-331.

Oulasvirta, A. & Blom, J. (2008). Motivations in personalisation behaviour. *Interacting with Computers,* 20:1, 1 -16.

Oviatt, S., MacEachern, M., & Levow, G. (1998). Predicting hyperarticulate speech during human-computer error resolution. *Speech Commun.* 24:2, May. 1998, 87-110

Paiva, A., Andersson, G., Höök, K., Mourão, D., Costa, M. & Martinho, C. (2002). SenToy in FantasyA: Designing an affective sympathetic interface to a computer game. *Personal Ubiquitous Comput*. 6:5-6, Jan, 378-389.

Parent, A. (2002). Jules Bernard Luys: A Singular Figure of 19th Century Neurology. *The Canadian Journal of Neurological Sciences,* 29:3, 282–288.

Partala, T. & Surakka, V. (2004). The effects of affective interventions in human-computer interaction. *Interacting with Computers*, 16, 295-309.

PCSpy Anytime (2009). *Waresight website.* Retrieved 20 September, 2009, from http://www.waresight.com/.

Penfield, W. & Rasmussen, T. (1950). *The Cerebral Cortex of Man*. New York: Macmillan.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A. & Booth, R. J. (2007). The development and psychometric properties of LIWC2007. Austin, TX: LIWC.net *Linguistic Enquiry and Word Count website*: Retrieved 20 September, 2009, from http://www.liwc.net/index.php.

Penton-Voak, I. S., Allen, T, Morrison, E. R., Gralewski, L. & Campbell, N. (2007). Performance on a face perception task is associated with empathy quotient scores, but not systemizing scores or participant sex *Personality and Individual Differences,* 43:8, Dec, 2229-2236.

Percy, L. (2001). The role of emotion in processing low involvement advertising. In: Groeppel-Klien, A. & Esch, F.R. (eds.) *European Advances in Consumer Research* 5. Provo, UT: Association for Consumer Research, 293-296.

Peter, C., Ebert E. & Beikirch, H. (2005). A wearable multi-sensor system for mobile acquisition of emotion-related physiological data. In: *Proceedings of the 1st International Conference on Affective Computing and Intelligent Interaction*, Beijing, 2005, Berlin, Heidelberg, New York: Springer Verlag, 691-698.

Peter, C. & Herbon, A. (2006). Emotion representation and physiology assignments in digital systems. *Interacting with Computers*, 18:2, 139-170.

Phillips, D. P. (1983). The impact of mass media violence on US homicides. *American Sociological Review,* 48, Aug, 560-568.

Picard, R. W. (1997). *Affective computing.* Cambridge, MA: The MIT Press.

Picard, R. (2009). Future Affective Technology for Autism and Emotion Communication. In: *Philosophical Transactions of the Royal Society, B.* Retrieved 20 September, 2009, from http://affect.media.mit.edu/pdfs/09.Picard-PhilTranRoyalSocB.pdf.

Picard, R. & Klein, J (2002). Computers that recognise and respond to user emotion: Theoretical and practical implications. *Interacting with Computers*, 14, 141-169.

Plato, translated by Jowett, B. (1999). *The Project Gutenberg ebook of Phaedrus*. Retrieved 10 May, 2010 from http://www.gutenberg.org/catalog/world/readfile?pageno=41&fk_files=920479.

Pleoworld. (2009). *Pleoworld website.* Retrieved 20 September, 2009, from http://www.pleoworld.com/Home.aspx.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In: Plutchik R. & Kellerman, H. (eds.) *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion.* New York: Academic, 3-33.

Plutchick, R. (2003). *Emotions and life.* Washington, D.C.: American Psychological Association.

Poh, M. Z., Swenson, N. C. & Picard, R. W. (2010). A wearable sensor for unobtrusive long term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 57:5, May 2010, 1243-1252.

Prance, R. J. S., Beardsmore-Rust, T., Watson, P., & Prance H. (2008). Remote detection of human electrophysiological signals using electric potential sensors. *Appl. Phys. Lett*., 93. Retrieved 20 September, 2009, from http://www.sussex.ac.uk/pei/documents/applab933033906_1.pdf.

Pronin, E., & Wegner, D. M. (2006). Manic thinking: Independent effects of thought speed and thought content on mood. *Psychological Science,* 17(9), 807-813.

Provine, R. (2000). *Laughter: A scientific investigation*. UK: Faber and Faber.

Qualifications and Curriculum Authority (2009). *National Curriculum 2009 Personal Wellbeing Key Stage 3* Retrieved 20 September, 2009, from http://curriculum.qcda.gov.uk/key-stages-3-and-4/subjects/pshe/personal-wellbeing/keystage3/index.aspx.

Reeves, B. & Nass, C. (1996). *The media equation: how people treat computers, television, and new media like real people and places.* Cambridge University Press.

Rehm M. & Wissner, M. (2005). Gamble - A multiuser game with an embodied conversational agent. In *Entertainment Computing,* ICEC 2005, 4th International Conference, 2005, 180-191.

Reiman, E. M., Lane, R. D., Ahern, G. L., Schwartz, G. E., Davidson, R. J., Friston, K. J., Yun L. S. & Chen, K. (1997). Neuroanatomical correlates of externally and internally generated human emotion. *Am J Psychiatry,* 154, 918-925.

Reisberg, D. & Heuer, F. (2004). Memory for emotional events. In Reisberg D. & Hertel P. (eds.) *Memory and emotion.* London: Oxford University Press, 3-41.

Reisenzein, R. (1995). On Oatley and Johnson-Laird’s theory of emotion and hierarchical structures in the affective lexicon. *Cognition & Emotion*. 9, 383-416.

Relational Agents website. Retrieved 20 September, 2009, from http://www.ccs.neu.edu/research/rag/

Revelle, W. & Loftus, D. 1992. The implications of arousal effects for the study of affect and memory. In: Christianson, S.A. (ed.) *Handbook of emotion and memory. Research and theory.* Hillsdale, NJ: Lawrence Erlebaum Associates. 113-149.

Ribi, F. N., Yokoyama, A. & Turner, D. C. (2008). Comparison of children's behavior toward Sony's robotic dog AIBO and a real dog: A pilot study. *Anthrozoos: A Multidisciplinary Jnl of The Interactions of People & Animals*, 21:3, Sep, 245-256.

Richards, J. M. & Gross, J. J. (2000). Emotion regulation and memory: the cognitive costs of keeping one's cool. *Jnl of personality and social psychology*, 79:3, 410-24.

Riek, L. D., Afzal, S. & Robinson, P. (2008). Affect decoding measures and human-computer interaction. In: *Proc. Measuring Behavior*. Maastricht, Aug 2008. Retrieved 20 September, 2009, from http://www.roboticschick.org/papers/FPS_usability_Riek.pdf.

Rinaldi, W. (1995). *The social use of language programme: Primary and infant school: Teaching pack number 1*. Chilworth, Surrey: Wendy Rinaldi.

Riseberg, J., Klein, J., Fernandez, R., & Picard, R. W. (1998). Frustrating the user on purpose: using biosignals in a pilot study to detect the user's emotional state. In: *CHI 98 Conference Summary on Human Factors in Computing Systems,* LA, California, United States, 18 – 23 April, 1998, CHI '98, ACM, New York, NY, 227-28.

Roberts, G. K. & Bavelas, J. B. (1996). The communicative dictionary: A collaborative theory of meaning. In: Stewart, J. (ed.) *Beyond the symbol model: Reflections on the nature of language.* Albany, NY: SUNY Press, 139-164.

Robinson, S., Hoheisel, B., Windischberger, C., Habel, U., Lanzenberger, R. & Moser, E. (2005). fMRI of the emotions: towards an improved understanding of amygdala function. *Current Medical Image Review,* CMIR, 2005; 1:2, 115-129.

Rojek, C. (2009). *The labour of leisure: The culture of free time*. Sage.

Rosenkranz, M. A. & Davidson, R. J. (2009). Affective neural circuitry and mind-body influences in asthma. *NeuroImage*, 47, 972–980.

Rosenthal, R. & Jacobson, L. (1992). *Pygmalion in the classroom*. New York: Irvington.

Rowland, B. A. & Steirn B. E. (2007). Multisensory integration produces an initial enhancement. *Front Integr Neurosci*. 1:4. Published online 2007 November 30. Retrieved 20 September, 2009, from http://frontiersin.org/neuroscience/profiles/barryestein/.

Rubin, D. C. & Kozin, M. (1984). Vivid Memories. *Cognition*, 16, 81–95.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology,* 39, 1161–1178.

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, 110:1, 145-172.

Ryan, T. P. (2007). *Modern experimental design*. Wiley.

Sackheim, H. A., Gur, R. C. & Saucy, M. (1978). Emotions are expressed more intensely on the left side of the face. *Science,* 202, 434-436.

Saffrey, C., Summerville, A., & Roese, N. J. (2008). Praise for regret: People value regret above other emotions. *Motivation and Emotion*, 32:1, 46-54.

Salovey, P. & Mayer, J. D. (1990). Emotional intelligence, *Imagination, Cognition and Personality*, 9, 185-211.

Satake, S., Kanda, T., Glas, D. F., Imai, M., Ishiguro, H., & Hagita, N. (2009). How to approach humans?: strategies for social robots to initiate interaction. In: *Proceedings of the 4th ACM/IEEE international Conference on Human Robot interaction,* La Jolla, California, USA, 09 – 13 March, 2009, HRI '09, ACM, New York, NY, 109-116.

Schacter, S., & Singer, J. E. (1962). Cognitive, social, and psychological determinants of emotional state. *Psychological Review*, 69, 379-399.

Schacter, D. (1996). *Searching for memory: the brain, the mind, and the past.* New York: Basic Books.

Scheirer J., Fernandez R., Klein J. & Picard R.W. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, 14:2, February 2002, 93-118.

Scherer, K. (1987). Towards a dynamic theory of emotion: The component process model of affective states. *Geneva Studies in Emotion and Communication*, 1:1, 1–98.

Scherer, K. R. (1995). Plato's legacy: Relationships between cognition, emotion and motivation. *Geneva Studies in Emotion and Communication*, 9:1, 1–7.

Scherer, K. R. (1999). Appraisal theory. In: Dalgleish T. & Power, M. (eds.) *Handbook of cognition and emotion.* John Wiley, New York, 637–663.

Scherer, K. (2003). Vocal communication of emotion: A review of research Paradigms. *Speech Commun*, 40, 227-256.

Scherer, K. (2004). Theories and models of emotion: A swamp (?) *Humaine Geneval Workshop WP3 Presentation.* Retrieved 20 September, 2009, from http://emotion-research.net/projects/humaine/ws/wp3/ExtraMaterial/HUMAINE-Scherer.pdf.

Scherer, K. R., Shorr, A. & Johnstone, T. (eds.) (2001). *Appraisal processes in emotion: theory, methods, research*. Canary, NC: Oxford University Press.

Scherer, K., Wranik, T., Tran, V., & Scherer, U. (2004). Emotions in everyday life. *Social Science Information,* 43:4, 499-570.

Schienle, A., Walter, B., Schäfer, A., Stark, R. & Vaitl, D. (2005). Gender differences in the processing of disgusting and fear-inducing pictures: an fMRI study. *NeuroReport*, 16, 277-280.

Schiffrin, D., Tannen, D. & Hamilton, H. (2001). *The handbook of discourse analysis*. Blackwell.

Schroeder, M. (2004). *Speech and Emotion research: an overview of research framework and a dimensional approach to emotional speech synthesis.* Doctoral Thesis, Phonus 7, Res Inst Phonetics, Saarland University Retrieved 20 September, 2009, from http://www.dfki.de/~schroed/articles/schroeder_phd_2004.pdf.

Schuller, B., Müller, R. Hörnler, B. Höthker, A. Konosu H. & Rigoll, G. (2007). Audiovisual recognition of spontaneous interest within conversations. In: *Proc. 9th Int. Conf. on Multimodal Interfaces (ICMI), Special Session on Multimodal Analysis of Human Spontaneous Behaviour,* Nagoya, Japan, ACM SIGCHI, 30-37.

Schuller, B., Muller, R., Lang, M., Rigoll, G., (2005). Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles. In: *Proceedings of Interspeech, 2005*, Lisbon, Portugal, 805-808.

Schulz, F. v. T. (1981). Miteinander Reden - Sto- erungen und Klaerungen. Reinbeck bei Hamburg: Rowolth Taschenbuch Verlag GmbH.

Schwartz, B. (2004). The tyranny of choice. Scientific American, April, 71-75.

SC-Keylog. (2009). *Soft Central website*, Retrieved 20 September, 2009, from http://www.soft-central.net/keylog.php.

Scollon, C. N., Diener, E., Oishi, S. & Biswas-Diener, R. (2004). Emotions across cultures and methods. *Journal of Cross-Cultural Psychology*, 35, 304-326.

Searle, J. (1969). *Speech acts*. Cambridge University Press.

Semaine website. Retrieved 20 September, 2009, from http://semaine-project.eu/publications

Sengers, P., Boehner, K., Mateas, M. & Gay, G. (2008). The disenchantment of affect. *Personal Ubiquitous Comput*., 12:5, Jun, 347-358.

Sengers, P., Kaye, J., Boehner, K., Fairbank, J., Gay, G., Medynskiy, Y. & Wyche, S. (2004). Culturally embedded computing. In: *IEEE Pervasive Computing*, 3:1, Jan.-Mar. 2004, 14-21.

Shan, T., Chen, S., Sanderson, C. & Lovell, B. C. (2007). Towards robust face recognition for intelligent-CCTV based surveillance using one gallery image. In: *IEEE Conference on Advanced Video and Signal Based Surveillance,* AVSS, 2007, 470-475.

Shaver, P., Schwartz, J., Kirson, D. & O'Connor, C. (1987). Emotion Knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52, 1060-1086.

Silani, G., Bird, G., Brindley, R.; Singer, T., Frith, C. & Frith, U. (2008). Levels of emotional awareness and autism: An fMRI study. *Social Neuroscience*, 3:2, June, 97 – 112.

Sloman, A. (2001). Beyond shallow models of emotion. *Cognitive Processing*, 2:1, 177-198.

Sloman, A. (2009) *Cognition and Affect Project website*. Retrieved 20 September, 2009, from http://www.cs.bham.ac.uk/~axs/cogaff.html.

Smith, P. (1992). *The emotional labour of nursing*, London: Macmillan.

Smith, C. & Lazarus, R. (1993). Appraisal components, core relational themes, and the emotions. *Cognition & Emotion*, 7, 233-269.

Spence, S. H. (2003). Social skills training with children and young people: Theory, evidence and practice. *Child and Adolescent Mental Health*, 8:2, 84-96.

SSRM Tek. (2009). Retrieved 20 September, 2009, from http://74.220.207.82/~northamp/index.php.

St. Jacques, P. L., Bessette-Symons, B. & Scabeza, R. (2009). Functional neuroimaging studies of aging and emotion: Fronto-amygdalar differences during emotional perception and episodic memory. *Journal of the International Neuropsychological Society*, 15, Cambridge University Press, 819-825.

Steinbrück, U., Schaumburg, H., Duda, S. & Krüger, T. (2002). A picture says more than a thousand words - photographs as trust builders in e-commerce websites. In: *Proceedings of Conference on Human Factors in Computing Systems CHI 2002*, *Extended Abstracts,* New York: ACM Press, 748-749.

Steiner, J. E. (1973). The gustofacial response: observation on normal and anencephalic newborn infants. *Symp Oral Sens Percept*., 4, 254-78.

Stephen, T. D. & Harrison, T. M. (1986). Assessing Communication Style: A New Measure. *American Journal of Family Therapy*, 14:3, Fall, 213-34.

Strack, F., Martin, L. & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, 54, 768-777.

Sundström, P., Ståhl, A. & Höök, K. (2005). eMoto – A user-centred approach to affective interaction, In: *Lecture Notes in Computer Science, Proceedings from ACII05*. Beijing, China: Springer. Retrieved 20 September, 2009, from

http://www.sics.se/~kia/papers/sundstrom.pdf.

Suslick, K. S. & Rakow, N. A. (2001). A Colorimetric nose: 'Smell-seeing'. In: Stetter J.R. & Pensrose, W.R. (eds.) *Artificial chemical sensing: Olfaction and the electronic nose.* Pennington, NJ: Electrochem. Soc., 8-14.

Takahashi, T., Bartneck, C. & Katagiri, Y. (2005). TelMeA – expressive avatars in asynchronous communications. *The International Journal of Human-Computer Studies,* IJHCS, 62:2, 193-209.

Tamagochi. (2009). *Tamagochi website.* Retrieved 20 September, 2009, from http://www.tamagotchieurope.com/EN/index.html.

Tannen, D. (1990). *You just don't understand: Women and men in conversation.* Ballantine.

Taylor**,** J. A. (1953). A personality scale of manifest anxiety**.** *Journal of Abnormal and Social Psychology***,** 48:2, April, 285-290.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In: Tuma, A. H. & Maser, J. D. (eds.) *Anxiety and the anxiety disorders.* Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, 681-706.

Ten Have, P. (1999). *Doing Conversation Analysis: A Practical Guide*. Thousand Oaks: Sage.

Thaler, R. H. & Sunstein, Cass R. (2009), *Nudge: Improving Decisions About Health, Wealth, and Happiness.* New York: Penguin.

Thayer, J. F. & Johnsen, B. H. (2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology,* 41, 243-246.

Tickner, N. (2004). Computer Rage - How Not To Destroy Your New High-Tech Gifts. *University of Maryland Newsdesk.* Retrieved 20 September, 2009, from http://www.newsdesk.umd.edu/sociss/release.cfm?ArticleID=1010.

Tractinsky, N. (2004). Tools over solutions? Comments on Interacting with Computers special issue on affective computing. *Interacting with Computers,* 16:4, 751-757.

TUC (2007). *Lies, damned lies and lie detectors: introducing lie detector tests for benefit claimants*. TUC commentary. Retrieved 20 September, 2009, from http://www.tuc.org.uk/welfare/tuc-13272-f0.pdf.

Tucker, D. M., Watson, R. T. & Heilman, K. M. (1977). Affective discrimination and evocation in patients with right parietal disease. *Neurology,* 27, 947-950.

Tutt, D., Shaukat, M., Fraser, M., Hindmarsh, J. & McCarthy, M. (2007). Can you see what I see? Developing coherent support ford distributed collaborative video analysis with MiMeG. In: *Third International Conference on e-Social Science.* June, 2007. Retrieved 20 September, 2009, from http://www.merc.ac.uk/sites/default/files/events/conference/2007/papers/paper168.pdf.

UK Data Archive website. Retrieved 20 September, 2009, from http://www.data-archive.ac.uk/

University Of Southern California (1997). 'Mugspot' Can Find A Face In The Crowd: Face-Recognition Software Prepares To Go To Work In The Streets. *ScienceDaily*, *University Of Southern California website*, 12 November. Retrieved 20 September, 2009, from http://www.sciencedaily.com/releases/1997/11/971112070100.htm.

Ushida, H., Hirayama, Y. & Nakajima, H. (1998). Emotion model for life-like agent and its evaluation. In: *Proceedings of the Fifteenth National/Tenth Conference on Artificial intelligence/innovative Applications of Artificial intelligence*, Madison, Wisconsin, United States, American Association for Artificial Intelligence, Menlo Park, CA, 62-69.

Van Gogh, V. (1889). *Letter to Theo van Gogh*. Saint-Rémy-de-Provence, Sunday, 14 or Monday, 15 July 1889. Retrieved 20 September, 2009, from http://www.vangoghletters.org/vg/letters/let790/letter.html.

Velten, E. (1968). A laboratory task for induction of mood states. *Behavior Research and Therapy*, 6, 473-482.

Verona, E., Patrick, C., Curtin, J. J., Bradley, M. M., & Lang, P. J. (2004). Psychopathy and physiological response to emotionally evocative sounds. *Journal of Abnormal Psychology*, 113, 99-108.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wade, T. D. & Kendler, K. S. (2000). The relationship between social support and major depression: cross-sectional, longitudinal, and genetic perspectives. *J Nerv Ment Dis*., May, 188:5, 251-8.

Wager, T. D. & Ochsner, K. N. (2005). Sex differences in the emotional brain. *Neuroreport*, 16:2, 85-87.

Wagner, H. L., Buck, R. & Winterbotham, M. (1993). *Journal of Nonverbal Behavior*, 17:1, 29-53.

Wallace, A. (1999). *The psychology of the internet*. New York: Cambridge University Press.

Wallis, P., Moore, R., Fagerberg, P., Cavazza, M. & Wilks, Y. (2006). Emotion in Human agent Interfaces: *COMPANIONS Consortium: State of the Art Papers,* 3 Jan 2006. Retrieved 20 September, 2009, from http://www.companions-project.org/downloads/Companions_SoA3_Emotion.pdf.

Wang, R. & Fang, B. (2008). Affective computing and biometrics based HCI surveillance system. In: *International Symposium on Information Science and Engineering 1*, ISISE 2008, 192-195.

Ward, R. D., Cahill, B., Marsden, P. H. & Johnson, C. A. (2002). Physiological Responses to HCI Events - what produces them and how detectable are they? In: *Procs. HCI 2002*, September, 2002, South Bank University, London, 90-93.

Ward, R. D. & Marsden, P. H. (2004). Affective Computing: problems, reactions and intentions. *Interacting with Computers,* 16:4, 707-713.

Watkins, C.; Daniels, L.; Jack, C.; Dickinson, H.; van den Broek, M. (2001). Accuracy of a single question in screening for depression in a cohort of patients after stroke: comparative study. *BMJ*, Nov, 323:1159.

Watson, D. A. & Tellegen A. (1985). Towards a consensual structure of mood. *Psychological Bulletin*, 98, 219 - 235.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scale. *Journal of Personality and Social Psychology*, 54, 1063-1070.

Westermann, R., Spies, K., Stahl, G., & Hesse, F.W. (1996). Relative effectiveness and validity of mood induction procedures. A meta-analysis. *European Journal of Social Psychology*, 26, 557-580.

Wierzbicka, A. (1999). *Emotions across languages and cultures: diversity and universals.* Cambridge, UK: Cambridge University Press.

Wieser, M. J, Muhlberger, A., Alpers, G. W., Macht, M., Ellgring, H. & Pauli, P. (2006). Emotion processing in Parkinson's disease: dissociation between early neuronal processing and explicit ratings. *Clin. Neurophysiol.,* 117, 94–102.

Williams, P. & Drolet, A. (2005). Age-related differences in responses to emotional advertisements. *Journal of Consumer Research*, 32, 343-354.

Winton, W. M., Putnam, L. E. & Krauss R. M. (1984). Facial and Autonomic Manifestations of the Dimensional Structure of Emotion. *Journal of Experimental Social Psychology.* 20, 195-216.

Wolfe, C. D. & Bell, M. A. (2004). Working memory and inhibitory control in early childhood: Contributions from electrophysiology, temperament, and language. *Developmental Psychobiology*, 44, 68-83.

Woods, D. & Fassnacht, C. (2007). *Transana v2.20.* . Madison, WI: The Board of Regents of the University of Wisconsin System. Retrieved 20 September, 2009, from http://www.transana.org/.

Woolf, V. (1922). Jacob's Room. Retrieved 20 September, 2009, from http://www.readprint.com/chapter-10641/Jacob-s-Room-Virginia-Woolf.

Young, C. (1996). *Emotions and emotional intelligence*. Cornell University.

Zajonc, R. B. (1980). Feeling and Thinking: Preferences Need No Inferences. *American Psychologist*, 35, 151-175.

Zelenski, J. M. & Larsen, R. J. (1999). Susceptibility to affect: A comparison of three personality taxonomies. *Journal of Personality*, 67, 761-791.

Zeng, Z., Pantic, M., Roisman, G. I. & Huang, T. S. (2007). A survey of affect recognition methods: audio, visual and spontaneous expressions. In: *Proceedings of the 9th international Conference on Multimodal interfaces*, ICMI '07, Nagoya, Aichi, Japan, 12 – 15 November, 2007, New York, NY: ACM, 126-133.

Zimmerman, E. & Salen, K. (2003). *Rules of play - game design fundamentals*. Cambridge: The MIT Press.