DO PEOPLE EMOTE WHILE ENGAGED IN HCI OFFICE SCENARIOS?

Mónica Sara Santos, Maja Pantic
Department of Computing, Imperial College London
180 Queen's Gate, London SW7 2AZ

ABSTRACT

Affect plays a key role in human interactions with computers. It influences critical aspects of success in life, such as learning, decision-making and social interactions. This paper describes an experiment to assess if, how, and when people emote while working with computers in office scenarios. To the best of our knowledge, this is the first study that addresses this question. For the experiment we continuously recorded 4 people in their work environment for a week, while they worked in their daily tasks with the computer. By the analysis of the recordings, we concluded people do emote while engaged in HCI office scenarios. Afterwards we conducted an online questionnaire where we showed 22 video sequences selected from the previous recordings, and asked the respondents what affective states were being displayed. The respondents clearly labelled 17 videos, 5 for each of the affective states of frustration/stress, fatigue, and interest/concentration, 1 for satisfaction/happiness and 1 for boredom.

KEYWORDS

Human-Computer Interaction, Affect, Spontaneous Affective States, Experimental Study.

1. INTRODUCTION

It has been argued that social and emotional intelligence are the parts of human intelligence that most influence aspects of success in life, especially in social interactions, learning, and adapting to what is important (Pantic et al. 2005). Emotions play a critical role in cognitive processes in humans, such as organisation of memory and perception (Bower 1981), learning (Dolan 2002), goals generation, evaluation and decision-making (Damasio 1994), and communication (Birdwhistell 1970; Ekman and Friesen 1975). These are all key factors that influence the way humans interact with computers (Reeves and Nass 1996).

In this paper we describe the experiment we conducted to assess if and how users emote while working with computers in office scenarios (i.e. sitting in front of the computer, working on their everyday tasks). This experiment was the first stage of a project that aims at researching multimodal, affect-sensitive HCI. Enabling computers to understand affect and adapt the interaction accordingly would likely make the communication more natural, efficient, and trustworthy (Pantic and Rothkrantz 2003). Approaching a more natural human-like interaction is a determining factor in the future of HCI (Maat and Pantic 2006) and is the main objective of our research.

Most of the research done on affective computing deals with a small set of posed displays of affect (Zeng et al. 2007), typically six basic emotions defined by Ekman and Friesen: anger, disgust, fear, happiness, sadness and surprise (Ekman and Friesen 1969). The six basic emotions are arguably not the most common emotions displayed in everyday interactions (el Kaliouby and Robinson 2005). In office-based HCI scenarios, due to the nature of the typically performed tasks, affective states such as confusion, frustration, understanding, fatigue, and satisfaction are expected to occur more frequently than the basic emotions (Maat and Pantic 2006).

Also, posed displays of affect are very different from spontaneous ones (Cohn and Schmidt 2003; Zeng et al. 2007), so systems developed to address posed emotions have difficulties in handling spontaneous displays of emotions (Pantic et al. 2005). As we are trying to deal with users' affective states in realistic scenarios, we are interested in studying spontaneous displays of affect.

Some experiments concerning affect displays in HCI environments have been done in the past. (Partala and Surakka 2004) describes an experiment conducted to investigate the psycho-physiological effects of positive and negative affective interventions in human-computer interactions. In their experiment, subjects were exposed to pre-programmed mouse delays, while trying to solve an interactive puzzle. After that, positive or negative interventions were provided via a speech synthesiser and the subjects' responses were recorded and analysed. Kapoor et al. conducted an experiment to assess user frustration, by asking a set of children to solve a computer version of the Towers of Hanoi puzzle. They analysed the subjects' non-verbal multimodal data, trying to predict when they were feeling frustrated (Kapoor et al. 2007). Branco et al. observed the spontaneous facial expressions subjects portray while trying to format a document in Microsoft Word, with the objective of identifying adverse event occurrences in the user interface (Branco et al. 2005).

Despite the experiments previously mentioned, to the best of our knowledge, no experiment like the one we conducted has been done before. We aim at understanding if, how, and why people emote while working in office scenarios. To assess that, we continuously recorded people in their everyday work environments, as they performed their daily tasks with the computer, for a whole work week. We not only recorded the subjects' faces, but also their screen, so we could determine if the display of an affective state was related to the interface or not.

Afterwards, we analysed the video recordings and saved the sequences where the subjects were displaying any kind of affective state caused by the interface or the interaction with the computer.

In order to label these sequences with an affective state, we conducted an online questionnaire, where we showed 22 video sequences and asked the respondents to indicate what affective states were being displayed by the subject in the video. Additionally we asked whether or not the respondents would like the computer to do something to try to help them, if they were the person displaying the affective states in the video.

We had responses from a broad variety of people – from several countries, of different computer expertise levels, and ranging over a wide age span.

The results show that people do emote while engaged in HCI office scenarios, and that the affective states they experience are, in fact, close to those suggested by (Maat and Pantic 2006). There was a clear agreement in the labelling of three non-basic affective states – fatigue, frustration/stress, and interest/concentration – of five videos for each affective state. Two other affective states – confusion and boredom – weren't as unanimously identified. The only video that the respondents labelled as satisfaction/happiness also showed a very high level of agreement (92,11%) for that affective state.

2. EXPERIMENT

We carried out the experiment in three phases: recording subjects in their work environment, analysing the recorded videos, and conducting an online questionnaire.

2.1 Recordings

In the first phase we individually recorded 4 subjects continuously for several full days, with the objective of assessing whether or not they emoted while working in their everyday tasks. In order to determine if a certain display of affect was related to interface issues, we also recorded the computer screen.

The recorded subjects were 2 females and 2 males, with ages between 26 and 37 years old. All subjects had different nationalities, namely Greek, Portuguese, Serbian and Spanish.

We used a webcam to record the subjects' faces. In one of the cases we used a mirror positioned next to the subject, showing a reflection of what was displayed on the screen. For this case, we have the mirror and the face of the subject in the same recording (Figure 1). For the other three subjects, we used a second webcam to record what was being shown on the screen. In the end, the recordings of the face and the screen were synchronised.

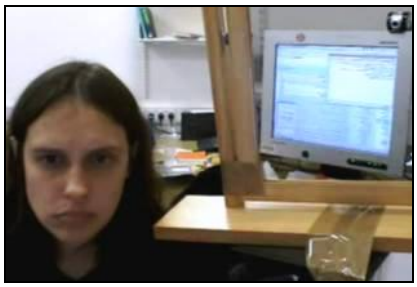


Figure 1. Illustration of the experiment setup. In this case the video shows both the subject and the screen (with the use of a mirror)

To record the data we used an independent computer, other than the subjects' own computers. We chose this option for two reasons: we didn't want the recording process to interfere with the subjects' work, for instance by slowing down the computer or overloading the local disk; and we didn't want to risk missing recordings in cases where the computer would crash, or the local computer processes would cause the recording itself to crash.

We recorded both video and sound. The videos were recorded at a 25fps frame rate and were compressed with the Xvid codec. In total 126 hours of footage were recorded.

In the second phase we analysed the recordings to verify whether or not people emoted. Whenever subjects showed any affective state, we extracted sequences into another file. The videos were viewed in fast forwarding mode and whenever the subject emoted, we rewound the recording and examined it in normal play mode. We examined both the portrayed affective states and the screen images, to determine whether or not the affective state had been caused by the interface. If so, the respective sequence was saved into another file.

2.2 Online Questionnaire

The third phase of the experiment had the objectives of labelling the previously selected recordings with the respective affective states, and assessing whether or not people would welcome assistance from the computer in certain situations.

In order to obtain responses from a variety of people we created an online questionnaire, so that it could be accessed and completed from any computer with access to the Internet. We were interested in responses from people with different levels of computer expertise, from different countries and cultural backgrounds, and within a wide age range. Respondents were recruited via email. Advertising emails were sent to university mailing-lists asking people to fill in the questionnaire. People were also encouraged to ask relatives and friends to respond, to ensure a balanced number of respondents from each expertise group. The whole recruiting / completing of questionnaire process lasted for a month.

The questionnaire was composed of two parts. The first part was intended to query respondents about their gender, age, profession, country of origin and level of expertise with the computer. All this data was important, as we wanted to analyse how different types of people responded to the same questions.

We defined three possible levels of expertise for the respondents to choose from: novice, intermediate and expert. We provided a short explanation of each level of expertise, so the respondents would know which level to choose. Novices were defined as being able to use the computer for everyday tasks, such as writing emails and visiting websites, but needing help from experts when faced with a different task or unexpected behaviour from the computer. The intermediate users were defined as those who use the computer regularly for a variety of tasks, and are able to solve most of the problems that occur, but might still need occasional help from experts. Finally, the experts were described as people who have training, or work in computing.

In the second part of the questionnaire we presented 22 video sequences selected from the recorded dataset. These selected sequences showed the subjects emoting. Even though we recorded both audio and video, we decided to use only video for this experiment. This decision was made for two reasons: on the one hand, since we were showing a large number of sequences, we wanted to keep their size as small as possible, so that they wouldn't take a long time to load. On the other hand, we wanted everyone who was labelling the

sequences to do so, based on the same cues. If we provided audio cues, some of the respondents might not be able to hear them, either for technical reasons or by preference.

For each of the sequences we asked:

1. What affective state(s) are displayed by the person in the video?
2. If you were the person in the video, would you like the computer to do something to try to help you?

A list of possible affective states was supplied for the first question, with the possibility of choosing multiple answers. We also provided an open answer text box, labelled "Other(s)". We decided not to have only open answers, as they are difficult to analyse statistically.

The decision on the list of affective states to provide as possible answers was based on two models of emotions: Russell's multidimensional emotional space (Russell 1980), and Plutchik's emotion wheel (Plutchik 1980). We decided to use some of the affective states displayed in each of the models, because neither of them covered all the affective states we were interested in. Although not being able to use only Russell's circumplex model, we mapped the selected affective states into it (Figure 2) and tried to have at least one affective state for each quadrant. The affective states we used for the questionnaire were satisfaction/happiness, interest/concentration, worriedness, inattention/distraction, surprise, confusion, fatigue, boredom, frustration/stress/anger, and shock. We also had a neutral option.

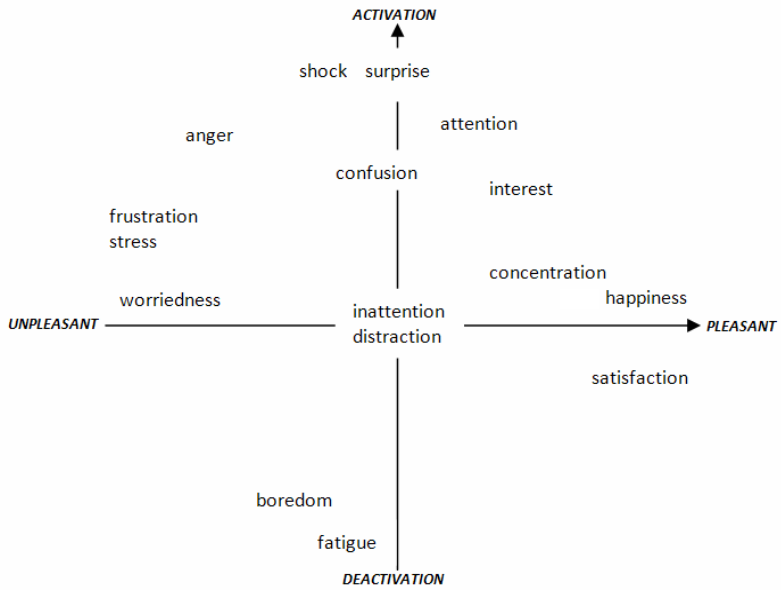


Figure 2. Mapping of the used set of affective states into Russell's circumplex model

3. RESULTS

75 people, comprising 38 females and 37 males, completed our online questionnaire. Given the ease of access the Internet provides, the replies arrived from people from 16 countries, with higher incidence of respondents from Portugal, The Netherlands, The United Kingdom, Greece and Brazil.

Figure 3 shows the age distribution of the respondents. The youngest respondent was 16 years old and the oldest 80 (both males). We had responses from 25 people of each level of computer expertise (novice, intermediate and expert).

Figure 4 shows the distribution of the respondents' gender by level of expertise. In the novice group there was a high incidence of women (72%), and in the expert group, the opposite happened – only 20% were females.

Figure 5 displays the absolute distribution of age groups of the respondents by their level of expertise. Interestingly the 30-39 years old age group is evenly represented in the three levels of expertise. As expected, groups of higher age are more represented in the lower levels of expertise. Only 1 respondent over 40 years old is an expert. The largest amount of experts belongs to the 23-29 years old group (15 people, which represent 63% of the total number of people in that age group). 67% of the people older than 60 years of age are novices.

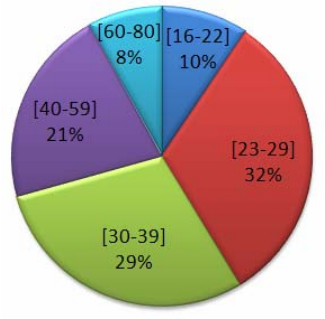


Figure 3. Age distribution of respondents

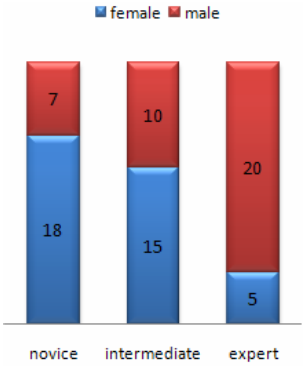


Figure 4. Distribution of respondents' gender by level of expertise

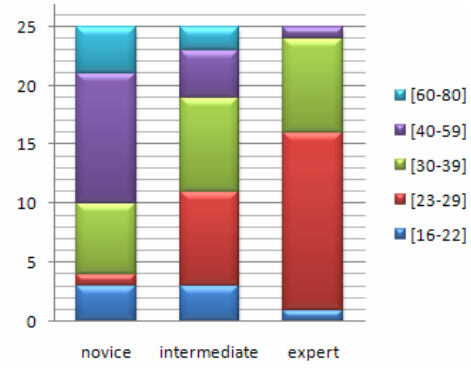


Figure 5. Distribution of respondents' age by level of expertise

Figure 6 shows the affective states' identification rate of all 22 videos, grouped by affective state. For the first three target affective states (fatigue, frustration/stress, and interest/concentration) we had a high level of agreement in labelling between respondents. These affective states were clearly identified in 5 out of 5 respective videos.

Of the first three affective states, fatigue was the one having the largest standard deviation. This happened because the identification of this affective state averaged 79,8%, but ranged broadly from 60,27% in one of the videos to 95,89% in another.

Frustration/stress was the affective state that was identified best, averaging 88,8%. Four of the frustration/stress videos were even labelled as showing frustration/stress by over 92% of the respondents. The other video was labelled with frustration/stress by 69,33% of the respondents.

Interest/concentration sequences were labelled as showing that affective state by an average of 74,74% of the respondents. Three of these videos were labelled as interest/concentration by over 80% of the respondents.

For the boredom sequences there was a much lower agreement ratio and it showed the highest standard deviation of all. From the three boredom videos, only one was obviously labelled as boredom, with 84,2% of agreement. The other two cases were identified by 46,58% and 55,26% of the respondents.

As shown in Figure 6, in the confusion case, there was no clear agreement in any of the videos. One of them ranked a little over 52% but the other two cases averaged much lower (around 40%), and the respondents seemed to be divided between confusion and surprise while labelling these videos.

Only one of the videos in the questionnaire was clearly labelled as satisfaction/happiness. This video was identified by 92,11% of the respondents as displaying that affective state.

We were also interested in knowing whether or not the recorded subjects would agree with the general opinion of the respondents, when it comes to the labelling of their own recordings. In order to know that, each of the four recorded subjects was asked to label their own videos, without knowing how those videos had been labelled by the respondents of the questionnaire. For two of the clearly identified affective states (two of the videos labelled as interest/concentration) the subjects gave a different answer from the majority of respondents. In both cases, the subjects labelled their own videos as showing confusion. For the rest of the cases, all subjects used the same label as the majority of the respondents.

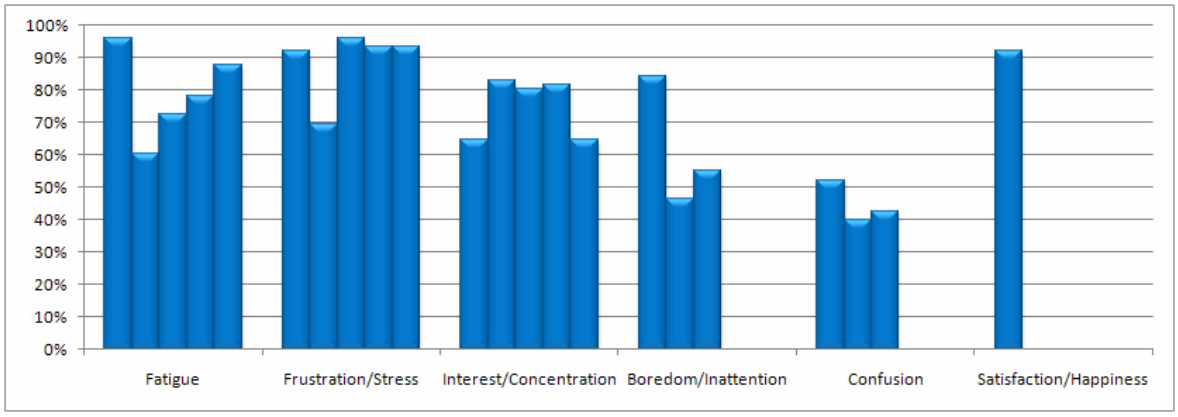
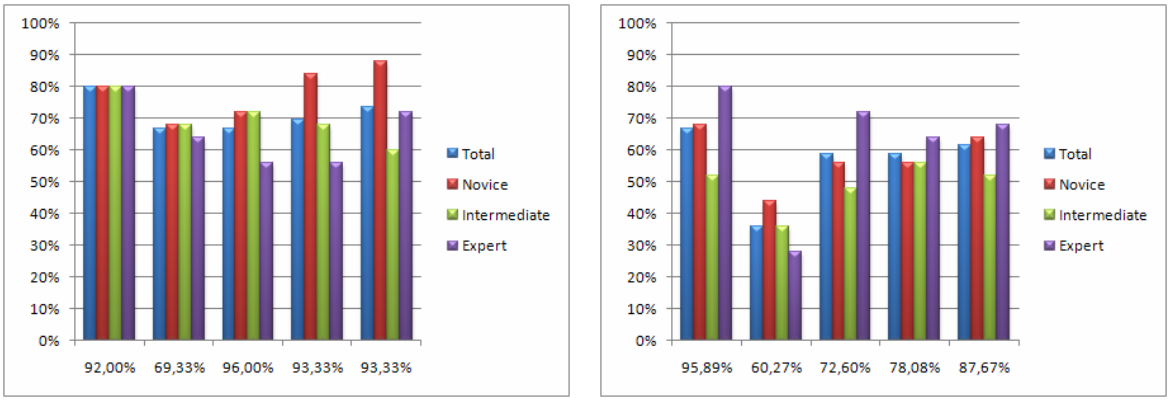


Figure 6. Affective states' identification rate per video

Another objective of our experiment was to assess whether or not people would welcome computer assistance, if they were the person in the video (this was question 2). Figure 7 shows the percentage of positive answers to this question for two affective states: frustration/stress and fatigue. We show the total percentage of positive answers and also the way people from each level of expertise responded to this question. In the horizontal axis we also display the percentage by which the respective video was labelled with frustration/stress (Figure 7.a) and fatigue (Figure 7.b).

The largest percentage of positive answers was given for the videos that were labelled with frustration/stress. For these videos, the lowest percentage of total positive answers was 66,7% and the highest 80%. In the fatigue cases, the percentage of positive answers was lower, but still above 50% in all videos but one, which happened to be the video that had a lower level of agreement on the affective state of fatigue.



a) **b)**

Figure 7. Percentage of respondents who replied *yes* to the question on whether or not they would welcome computer assistance, if they were the person in the video; the first bar represents the average % of respondents, the second the % of novices, the third one the % of intermediates and the last bar, the % of experts; bellow each set of bars is the % by which the video in question was labelled with the respective affective state; a) shows the 5 videos labelled with frustration/stress; b) shows the 5 videos labelled with fatigue

Given their more positive nature, interest/concentration videos had lower numbers of positive answers to question 2, not reaching 50% for any of the videos. On the satisfaction/happiness video, 17,3% of the respondents said they would like the computer to try to help them. For the remaining videos, even though they couldn't be clearly identified with an affective state, the percentage of positive answers was over 50%.

4. DISCUSSION OF RESULTS

Even though we decided to leave audio cues out of the online questionnaire and show only videos without sound, we had very satisfactory results in the identification of non-basic affective states.

The three affective states of fatigue, frustration/stress and interest/concentration had high rates of agreement from respondents. This sets a baseline of what are the possible behavioural signals that identify these affective states and we can do further work from this point.

For boredom, only one of the video sequences yielded good identification results, and for confusion none of the videos could be unambiguously labelled as displaying that affective state. Confusion seems to be a difficult affective state to identify. The best of our results showed an agreement rate of only 52%. In the other two videos of this affective state, there were a significant number of respondents who labelled the videos with surprise, instead of confusion.

Hence, for these two cases (boredom and confusion) we will have to step back and try to find another way of studying them, perhaps by finding other video sequences that might be more easily labelled with those affective states.

When it comes to the respondents' answers to the question on whether or not they would like the computer to try to help them, if they were the person in each video (question 2), we, as expected, had the largest percentage of positive answers in the videos that showed negative emotions. Frustration/stress videos were the ones that yielded more positive responses to this question. For those videos, the same or a larger amount of novices, when compared to intermediate and experts, responded positively to this question.

In the fatigue cases, the percentage of positive answers was lower than in the frustration/stress ones, but still above 50% in all videos but one, which happened to be the video that had a lower level of agreement on the affective state of fatigue.

An interesting fact is that for 4 out of the 5 fatigue videos, unlike what happened in the frustration/stress case, experts displayed a higher percentage of positive answers than the other expertise groups.

The results of the frustration videos aren't surprising, as novices, due to their lack of experience with the computer, are usually prone to having any possible kind of help while working with it. However, the results of this same question when it comes to fatigue provided unexpected results. We were expecting the experts to seldom answer positively to question 2, and for almost all of the videos, they are the group that has lower rates of positive answers to this question. The videos displaying fatigue were an exception: for this affective state, experts are actually more open to computer assistance than those in other expertise categories.

In most of the videos that weren't clearly labelled with one affective state there were still over 50% positive answers to question 2. This seems to indicate that, for these cases, although not being able to agree on an affective state, people still identified the need for an intervention from the computer.

5. CONCLUSIONS

This paper described the experiment we conducted to assess whether or not people emote while working in HCI office scenarios. We are interested in studying spontaneous affective states that occur in human interactions with the computer. In the first stage of the experiment we continuously recorded 4 subjects for a week whilst they worked with a computer in their natural working environment.

By analysing the videos, we concluded that people do emote while engaged in HCI office scenarios. After cutting sequences where the subjects were emoting, we conducted an online questionnaire where we asked people to label the selected videos with the affective states being displayed. We additionally asked them to state if they would welcome assistance from the computer, if they were the person in the video (question 2).

Only using video cues, for the fatigue, frustration/stress, and interest/concentration video sequences, the experiment yielded very positive results, as those affective states were clearly identified by the respondents of the questionnaire. Given this level of agreement, we can now proceed to the next step of the study, when it comes to these three affective states. This comprehends the clear definition of behavioural cues that are typical for each affective state, and the building of a tool that automatically detects these cues.

We will record other subjects while they work, and later save the sequences where they are emoting, so we can have more subjects showing spontaneous displays of our target affective states. After the selection of relevant sequences, we will do a manual annotation of cues that are common to each affective state. We will

analyse facial expression, head and hands movements, and sound – non-linguistic vocalisations (yawns, laughter, etc.) and swear words. After the identification of those behavioural cues, we'll build a system to automatically detect them, with the aim of creating an affect-sensitive human-computer interface.

We also had interesting results when it comes to the answers to question 2. For most of the videos shown, more than 50% of the people responded *yes* to this question. Most of the cases where the *no* answer prevailed were for the interest/concentration and the satisfaction/happiness videos. This was an expected result, given the positive nature of these affective states.

For all but one of the fatigue videos more expert respondents than novices or intermediates responded *yes* to question 2. This shows that there are cases where even experts feel like they would like the computer to interfere and execute some assisting action.

At a later stage we will study what would be appropriate reactions/feedback from the computer, while responding to users' affective states and tasks being performed.

ACKNOWLEDGEMENT

We would like to thank everyone who participated in the experiment, namely those who agreed to be filmed by us and those who completed our online questionnaire.

REFERENCES

Birdwhistell, R. L. (1970), *Kinesics and Context: Essays on Body Motion Communication* (University of Pennsylvania Press).

Bower, G. H. (1981), 'Mood and memory', *American Psychologist,* 36 (2), 129-48.

Branco, P., et al. (2005), 'Faces of emotion in human-computer interaction', *Conference on Human Factors in Computing Systems*, 1236-39.

Cohn, J. F. and Schmidt, K. (2003), 'The Timing of Facial Motion in Posed and Spontaneous Smiles', *Active Media Technology*.

Damasio, A. R. (1994), *Descartes' Error: Emotion, Reason and the Human Brain* (Picador).

Dolan, R. J. (2002), 'Emotion, Cognition, and Behavior', *Science,* 298 (5596), 1191-94.

Ekman, P. and Friesen, W.V. (1969), *The repertoire of nonverbal behavior* (Mouton de Gruyter).

Ekman, P. and Friesen, W. V. (1975), *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues* (Prentice-Hall).

el Kaliouby, R. and Robinson, P. (2005), 'Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures', *Real-Time Vision for Human-Computer Interaction*, 181–200.

Kapoor, A., Burleson, W., and Picard, R. W. (2007), 'Automatic prediction of frustration', *International Journal of Human-Computer Studies,* 65 (8), 724-36.

Maat, L. and Pantic, M. (2006), 'Gaze-X: adaptive, affective, multimodal interface for single-user office scenarios', *ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces*, 171-78.

Pantic, M. and Rothkrantz, L.J.M. (2003), 'Toward an affect-sensitive multimodal human-computer interaction', *Proceedings of the IEEE,* 91 (9), 1370-90.

Pantic, M., et al. (2005), 'Affective multimodal human-computer interaction', *Proceedings of the 13th annual ACM international conference on Multimedia*, 669-76.

Partala, T. and Surakka, V. (2004), 'The effects of affective interventions in human–computer interaction', *Interacting with Computers,* 16 (2), 295-309.

Plutchik, R. (1980), *Emotion: A Psychoevolutionary Synthesis* (Harper & Row New York:).

Reeves, B. and Nass, C. (1996), *The media equation: how people treat computers, television, and new media like real people and places* (Cambridge University Press New York, NY, USA).

Russell, J. A. (1980), 'A circumplex model of affect', *Journal of Personality and Social Psychology,* 39 (6), 1161-78.

Zeng, Zhihong, et al. (2007), 'A survey of affect recognition methods: audio, visual and spontaneous expressions', *ICMI '07: Proceedings of the 9th international conference on Multimodal interfaces*, 126-33.