Available online at www.sciencedirect.com
SCIENCE DIRECT®

Interacting with Computers 18 (2006) 227–245

Interacting with Computers

www.elsevier.com/locate/intcom

Empathic agents to reduce user frustration: The effects of varying agent characteristics

Kate Hone*

School of Information Systems, Computing and Mathematics, Brunel University, St Johns Building, Kingston, Uxbridge UB8 3PH, UK

Received 12 October 2004; revised 5 May 2005; accepted 15 May 2005
Available online 4 January 2006

Abstract

There is now growing interest in the development of computer systems which respond to users' emotion and affect. We report three small-scale studies (with a total of 42 participants), which investigate the extent to which affective agents, using strategies derived from human–human interaction, can reduce user frustration within human–computer interaction. The results confirm the previous findings of Klein et al. [Klein, J., Moon, Y., Picard, R.W., 2002. This computer responds to user frustration: theory, design and results. Interacting with Computers 142, 119–140] that such interventions can be effective. We also obtained results that suggest that embodied agents can be more effective at reducing frustration than non-embodied agents, and that female embodied agents may be more effective than male embodied agents. These results are discussed in light of the existing research literature.

Keywords: Affective computing; User emotion; Frustration; Agents

1. Introduction

'Affective computing' can be defined as 'computing that relates to, arises from, or deliberately influences emotion' (Picard, 1997). A number of different types of research are encompassed within this term. For instance, some Artificial Intelligence researchers in the field of affective computing are interested in how emotion contributes to human

* Tel.: +44 1895 266009; fax: +44 1895 251686.
E-mail address: kate.hone@brunel.ac.uk.

0953-5438/$ - see front matter
doi:10.1016/j.intcom.2005.05.003

and, by analogy, computer problem solving or decision making (e.g. Belavkin, 2001). Others are concerned with enabling human–human communication of emotion through the medium of computer networks. Example applications here include 'expression glasses' which allow any viewer to visualise the confusion and interest levels of the wearer (Scheirer et al., 1999) and affective avatars where a full-face mask on the user drives a representation on the avatar's computer-graphic face (see Picard, 1997). Underlying many of these different strands of research is work on understanding the nature of emotion and how it should be represented (e.g. Schiano et al., 2000). The area of affective computing research with which this paper is primarily concerned is the design of systems that respond to user emotion with the aim of improving human–computer interaction. Specifically, we look at systems that respond to user frustration at the interface and use interventions to try to dissipate that frustration. The studies described here build upon previous research on this topic by Klein et al. (2002) from MIT.

The paper proceeds as follows. We begin by briefly describing research that is currently being conducted around the world to enable computers of the future to recognise user emotion through channels such as face and voice recognition. This demonstrates the vibrancy of this research area and suggests that it may be technically feasible for computers to detect user emotional states during human–computer interaction. We then go on to discuss potential applications of this technology, specifically considering how computers might respond once they have detected a particular user emotion within a particular interaction context. We then describe Klein et al. (2002) work on computers that respond to user frustration. This leads to a discussion of the research questions explored within the current research. There then follows an account of three small-scale experiments, which investigate these questions. We end with a discussion of the findings, their limitations and further work which is required in this area.

2. Background

2.1. Automatic detection of human affect and emotion

Humans naturally use a number of visible and audible cues in order to recognise emotional states in other people. Facial expression is particularly important in the communication of emotion and there is evidence for the existence of a number of universally recognised facial expressions for emotion, namely happiness, surprise, fear, sadness, anger and disgust (Ekman, 1982). The body (gesture and posture) and tone of voice are the other main channels for the communication of emotion (Argyle, 1988). There are also a number of psycho-physiological correlates of emotion (e.g. pulse or respiration rate), most of which cannot easily be detected by human observers, but which could be made available to computers given appropriate sensing equipment. The sections which follow briefly review some of the technological developments which are taking place to allow such emotion signals to be automatically detected and recognised by machine.

2.1.1. Facial expression recognition

Several research groups are investigating the automatic decoding of emotion from facial expression. For example, Jeff Cohn and his colleagues at the University of

Pittsburgh are working on methods for discriminating emotion from facial images (e.g. Cohn and Kinade, in Press). Their approach uses computer vision and artificial neural networks and is based on the Facial Action Coding System (FACS) developed by Ekman and Friesen (1978). FACS is one of the most highly developed methods for coding facial expression, based on the analysis of small facial movements which are visible to human observers and discriminable from each other (Argyle, 1988). Results with the automated system using test images are promising, with evidence of high concurrent validity with human observers. A number of other researchers are working on neural network approaches or knowledge-based approaches to facial expression recognition (Lisetti and Schiano, 2000). While considerable progress has been made in recent years, the systems remain experimental and only work with carefully prepared images rather than real-time video data. Where attempts have been made to use video data, human subjects are required to place their faces in a vice at a set distance from the camera, which is clearly impractical for most real world applications.

2.1.2. Recognition of emotional expression in the voice

A number of research groups are investigating the decoding of emotion from speech signals. An example is the work at ATR Research Laboratories in Japan (e.g. Nakatsu et al., 1999). Their approach involves developing a database of speech samples where speakers are asked to express one of the following emotions: anger, sadness, happiness, fear, surprise, disgust, playfulness and neutrality. The speech signals are then analysed to extract phonetic and prosodic (amplitude, temporal structure and pitch) features and this data is used to train an artificial neural network. Testing (with samples generated in the same way as the training data) produced results comparable to human listeners. However, it is unclear whether these results would generalise to naturally produced, rather than 'acted', emotional expression. Nakatsu et al. (1999) describe an implementation of their technology in an interactive movie system. They combined traditional speech recognition, emotion from voice recognition and gesture recognition (using special wearable sensors), using the inputs to alter the course of a movie scenario. However, they have not yet evaluated the effectiveness of the emotion recognition component in this context and the system currently only works for male speakers.

2.1.3. Kinaesthetic detection of emotion state

Another approach to emotion recognition is to assess psycho-physiological changes in the user. Researchers at the MIT Media lab have been using sensors which detect galvanic skin response (GSR), blood volume pulse, respiration rate and electromyographic activity of muscles (e.g. Picard, 1998). The 'Emotion Mouse' developed by IBM detects pulse rate, GSR, skin temperature and general somatic activity and uses this data to categorise user emotional state (Ark et al., 1999). They claim good results with this set-up, though this conclusion was based on a study of only six people and used 'acted out' rather than real emotion.

2.2. The role of emotion in human–computer interaction

Most researchers working on automatic emotion detection hope that their work will eventually lead to improved human–computer interaction. Picard (1997) for instance, maintains that giving computers the ability to recognise, react to and express emotions will make them more effective at communicating with their human users. Often the argument presented appears to rest on two premises. First, that emotional recognition and expression are important for human–human communication. Second, that humans behave towards computers as they do towards other humans. This second argument is based primarily on the controversial findings of Reeves and Nass (1996) who, in a series of empirical studies, concluded that humans respond socially and naturally to media (including computer interfaces) and that these responses are unconscious. An example of one of their findings is that people give more honest assessments of the quality of a computer program when asked by a different computer from the one on which they used that computer program, mirroring the pattern shown in human–human communication. While the work of Reeves and Nass (1996) is frequently cited by researchers in the field of affective computing, none of the experiments which they conducted pertain directly to the question of whether systems which recognise human emotion can improve human–computer interaction. At best their work suggests that users may unconsciously alter their emotional expressions (facial and voice) when communicating with a computer, since this is natural, social behaviour. It is worth noting that not all human emotional signals are unconscious, in fact it is very common for humans to deliberately send emotional expressions which do not necessarily reflect their inner state (Argyle, 1988). It is, therefore, also possible that users will learn to use emotional expressions in their interactions with computers if they see a benefit in doing so.

Assuming that users will provide appropriate cues to their emotional state that systems can recognise, the next question is how that information might be used in shaping the human–computer interaction. A number of applications have been proposed which might benefit from emotion recognition components, including:

- computer aided learning systems which change the pace or content of a computer-based tutorial based on sensing the level of interest or puzzlement of the user (e.g. Picard, 1997; Lisetti and Schiano, 2000),
- ubiquitous computing applications such as an 'intelligent' CD player which selects music on the basis of the user's affective state (Healey et al., 1998),
- entertainment applications such as games or interactive movies where the action changes based on the emotional response of the user (e.g. Nakatsu et al., 1999),
- help systems which detect frustration or confusion and offer appropriate user feedback (e.g. Klein et al., 2002).

This last application area is the focus of the current paper and the section which follows, therefore, considers this topic in more detail.

2.3. Computers that respond to user frustration

Klein et al. (2002) designed and tested an intervention intended to relieve user frustration caused by a computer application. This intervention consisted of a text-based interaction with a computer 'agent' using a dialogue strategy known to be effective at relieving negative emotion in human–human interactions. In an experimental trial, participants receiving the frustration relieving intervention continued to interact with the computer which had frustrated them, for longer than participants receiving a control intervention. This result is taken as support for the view that the agent was successful in undoing some of the negative feeling caused by the computer. While this result is encouraging it should be noted that the researchers were unable to find evidence of frustration reduction in the participants' subjective ratings and the relevance of subsequent interaction time as a measure of reduced frustration may be questionable. Another limitation of the study was that the agent's intervention followed a set period of interaction with the frustrating application rather than being driven by real-time detection of frustration. The research reported in this paper attempts to overcome these limitations and to explore further the effects of agent characteristics on the degree of frustration reduction achieved.

Klein et al. (2002) used a text-based agent designed with the intention of reducing user frustration. The agent used 'active listening', empathy and sympathy with the intention of helping to relieve the negative state induced by a frustrating computer program. These techniques were chosen because they have been shown to be effective in human–human communication situations (such as therapy). In the experimental trials, frustration was induced by a game in which participants experienced nine seemingly random 'web delays' (control conditions were also run where users did not experience any delay). The frustration reducing agent was compared to two other interventions: ignore and vent. In each case users interacted with the game for 5 min after which a text-box popped up on screen. In all conditions users were asked some general demographic questions. In the ignore condition participants were then asked to rate the game using purely factual criteria.

In the agent and vent conditions participants were asked to rate the game with a series of emotional-type responses. They were then asked to rate their level of frustration on a 10-point scale. In the agent condition what followed was an agent response tailored to the level of frustration expressed by the participant. For example, if the participant rated their frustration level as 7 out of 10 then the agent feedback would be "Hmmm. It sounds like you felt really frustrated playing this game. Is that about right?" (yes/no). There was also an expression of sympathy (for instance "It sounds like you didn't have the best experience, then. That's not much fun") and an empathy statement (e.g. "Sorry to hear things didn't go so well"). In the vent condition, the frustration rating was followed by a series of other emotional ratings and open ended questions which allowed participants to 'vent' their frustration at the machine.

Following the experience with either the agent or one of the other interventions, participants were asked to interact with the game again (this time with no delays). After a period of 3 min a 'quit' button appeared on the screen but participants could play for longer if they wished. After completing the experiment participants were asked to rate (on a paper questionnaire) how they felt at various points during the experiment. The main dependent measure used in analysing the results of this experiment was the time that participants spent interacting with the game during the second phase of the experiment. The prediction was that if frustration was effectively relieved, then participants would feel more positive towards the task and would, therefore, continue to interact with it for longer. The fact that participants in the delay conditions spent significantly longer interacting with the game after the agent intervention (compared the ignore and vent interventions) is, therefore, taken as evidence to support the efficacy of the agent's approach. However, this interpretation seems somewhat problematic, especially since there was a main effect of delay condition on time spent interacting, with those in the delay conditions interacting longer than those in the no-delay conditions. The authors' interpretation of the behavioural results might lead one to predict that those who feel better overall with the system (those experiencing no delay) would interact with the system longer.

Analysis of the self-report measures of frustration and emotional state did not reveal any significant effects. The authors argue that this is "consistent with…emotion theorists, who [argue] that self-report data tends to be unreliable" (Klein et al., 2002, p. 136). However, the use of this argument seems problematic given that the agent's feedback is itself tailored to the intensity of self-reported ratings of emotion. If these are thought to be unreliable it calls into question the theoretical basis for the agent design. The key problem may have been that the majority of the self-report data in Klein et al. (2002) was collected some time after the frustration had actually been experienced and participants' *recall* of their emotional state may not have been reliable. Interestingly, the data on emotional state collected on-line during the experiment as part of the agent and vent conditions did demonstrate significant differences in frustration between the delay and no-delay conditions. This confirms the effectiveness of the experimental manipulation of frustration, but also suggests that the real time measurement of frustration is capturing meaningful data.

Overall, while the research reported in Klein et al. (2002) presents a novel and potentially exciting means of reducing user frustration with computers, we argue that the empirical results are ambiguous as to whether this was in fact achieved. The research reported here sought to obtain clearer evidence for the efficacy of frustration-reducing affective agents. This work is reported in study 1 below.

A further limitation of the Klein et al. (2002) study was the very limited nature of the 'agent' used in the trials. The agent's behaviour consisted simply of text-boxes in which the affective feedback to users was presented. In contrast much contemporary research tends to focus on the use of animated interface agents (see Dehn and Van Mulder, 2000, for a review). A number of authors argue that by rendering computers more 'human-like', animated agents will increase user engagement and motivation, and consequently will improve human–computer interaction (Dehn and Van Mulder, 2000). However, evidence for such an improvement, in terms of either user attitude or performance, is somewhat equivocal from the available empirical evidence. Dehn and Van Mulder (2000) make the point that the effectiveness of an animated agent is likely to depend on the context (e.g. a game environment vs. a business environment), the characteristics of the agent used and the fit between these two elements. There are likely to be some application types where it is appropriate to have a human-like interaction and others where it is not. In the case of responding to user frustration in the manner proposed by Klein et al. (2002) it is worth noting that the interaction style used is wholly based upon strategies used in human communication. Given that this is the case it seems reasonable to hypothesise that this is an application where an animated agent might be advantageous, since the human-like message would have a good fit with the human-like appearance of an animated agent delivering it. One might, therefore, expect the effectiveness of affective feedback to be enhanced if the message appears to come from a human-like source. The second study reported here investigates this hypothesis. The final study reported here goes on to investigate the effect of varying the characteristics of the agent used to deliver the affective message.

3. Study 1: a partial replication of Klein et al. (2002)

3.1. Introduction

We chose to perform a much simplified version of Klein et al. (2002) experiment, addressing some of the shortcomings described above. A key difference between our work and Klein et al. (2002) was that frustration levels were measured during the experimental task itself. A first measure of frustration was obtained during the participant's interaction with the affective agent. This measure of frustration was used to tailor the agent feedback (as in Klein et al., 2002), but also as a baseline against which to compare frustration after the agent interaction. A second measure of frustration was taken immediately after the experiment, once the participant had interacted with the agent and then with the frustrating game for a second time.

A second way in which the experiment differed from Klein et al. (2002) study is that we also allowed participants to choose when to interact with the affective agent. Thus there was an on-screen icon which participants could press if they began to feel frustrated. We hoped this would represent a better simulation of future affective agents, which will only appear once they detect that the user looks frustrated.

As in Klein et al. (2002) study, we compared the affective condition to a non-affective control intervention. In this control condition users were asked to rate their level of frustration with the system, but did not subsequently receive affective feedback linked to this level of frustration. Unlike Klein et al. (2002), however, we did not include a non-frustrating game as a further control condition, since our main hypotheses only concern the effect of the agents when a user is frustrated. The aim of the experiment was to investigate whether the presence of an affective agent would lead to reductions in self-reported levels of frustration. The experimental hypotheses were, therefore, that:

- H1: participants interacting with the affective agent will experience significant reductions in their self-rated frustration levels,
- H2: participants interacting with the affective agent will experience greater reductions in self-rated frustration levels than those in the control condition.

3.2. Method

3.2.1. Procedure and experimental design

A between-groups experimental design was used. In the experimental condition, participants interacted with an affective agent with similar behaviour to that in Klein et al. (2002). This agent provided affective feedback to the user, based on that user's self-rating of their frustration level. The agent in the control condition was identical to the affective agent except that instead of using active listening and empathy it simply confirmed the user's selection (e.g. "you have selected number 2. Thanks for your time"). In both conditions participants interacted with a deliberately frustrating computer game.

The frustrating game created for this experiment was a simple, mouse operated bat and ball game where the player is in control of a paddle with which he/she deflects a ball around the playing area (Fig. 1). A point is scored with each deflection of the ball from the top of the playing area. Players were given four balls at the start of each game and a ball was lost when it was missed by the paddle. The ball was programmed to travel at random speeds and in randomly selected directions. In addition, the ball sometimes passed through the paddle and the paddle response was sometimes lagged, creating the impression of non-responsiveness due to a slow web connection. Participants were offered an incentive to beat a high score of 158 (a virtually unattainable level), which they were told had been achieved by a 16-year-old boy. This manipulation was intended to further increase the level of frustration experienced.

Participants were instructed that if they felt frustrated during the game they should click on an agent icon. This brought up either the affective agent (experimental condition) or the control interface (control condition). In each case the participant was asked to rate their level of frustration on a 5-point scale (where 5 is most frustrated). In the affective agent condition this led to tailored feedback (as in Klein et al., 2002). In the control condition, the user's selection was simply confirmed. The participants then resumed interaction with the game (which continued to behave erratically) for as long as they wanted to. Once they had finished they were asked to click on a second agent icon, which recorded their current level of frustration (again on a five point scale). The dependent variables in this experiment were

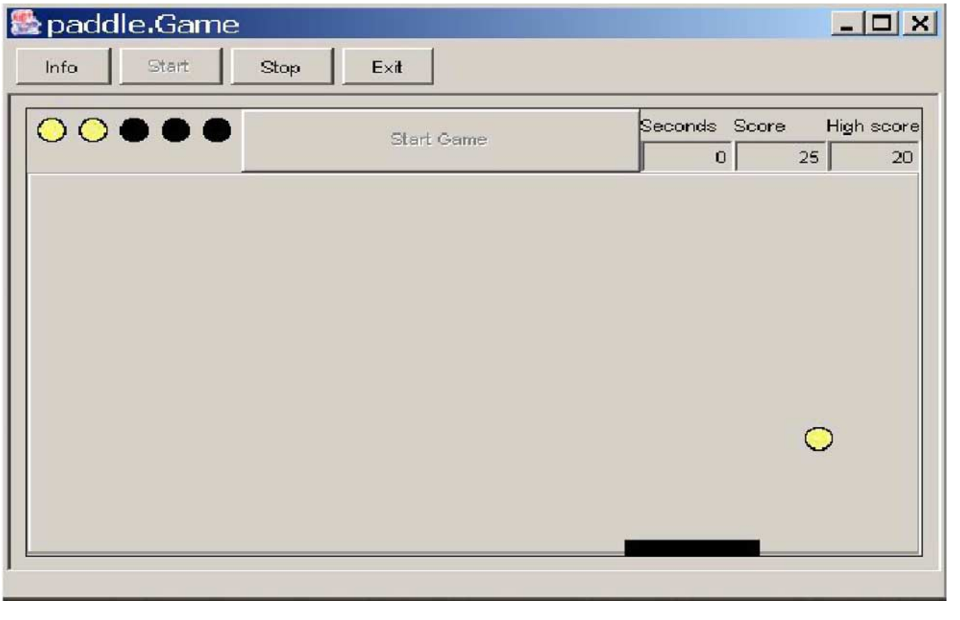

Fig. 1. Screen shot of frustration inducing game.

the frustration rating given in the first interaction with the system (time 1) and the frustration rating given at the end of the second period of interaction with the system (time 2).

3.2.2. Participants

Twelve participants took part in the study, with six assigned to each condition. All were students and were experienced users of computer systems. Participants were typically in the age groups 10–25. Experimental groups were balanced for gender, with three male and three female participants in each condition.

3.3. Results

Participants interacted with the frustrating game for an average of 5 min 26 s before choosing to interact with the agent. The frustration rating results are shown below in Fig. 2.

In the affective agent condition, the mean frustration rating before the agent feedback (time 1) was 3.33; the mean frustration rating after the intervention (time 2) was 2.0. Due to the small sample size, non-parametric statistical tests were used to explore the experimental hypotheses. A Wilcoxon signed-ranks test was used to compare the frustration ratings at times 1 and 2. Since, we hypothesised that frustration would be lower following the agent's intervention, a one-tailed test was used. The result showed a significant difference between the ratings at times 1 and 2 ($W=0$, $n=6$, $P<0.05$). We were, therefore, able to accept H1 that participants interacting with the affective agent will experience significant reductions in their self-rated frustration levels.

In the control condition the mean rating of frustration at time 1 was 2.83 and at time 2 was 3.17. A Wilcoxon signed-ranks tests revealed no significant difference between these scores.

In order to test H2 a Man–Whitney test was used to compare the change in frustration in the affective agent condition with the change in frustration in the control condition. The mean decrease in frustration in the affective agent condition was 1.33. In the control condition, frustration increased by an average of 0.5. The difference between the conditions was found to be significant ($U(6,6)=5.5$, $P<0.05$). We were, therefore, able to accept H2 that participants interacting with the affective agent will experience greater reductions in self-rated frustration levels than those in the control condition.

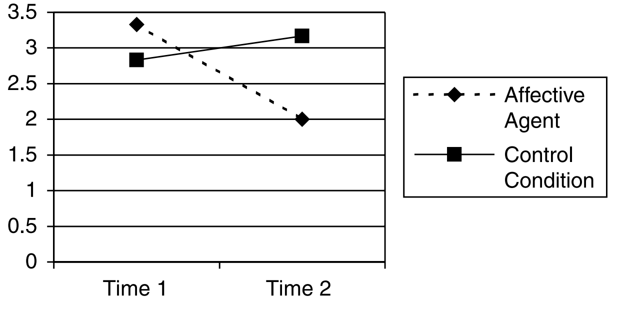

Fig. 2. Frustration scores from study 1 (frustration measured on a 5-point scale where 5 is most frustrated).

3.4. Conclusion

This experiment has demonstrated that an affective agent, using techniques of active listening and empathy, can reduce the level of frustration experienced by users of a frustrating computer application. It builds on the research reported in Klein et al. (2002) as it was able to demonstrate an effect on participants' self-reports of frustration level.

4. Study 2: text vs. embodied agent

4.1. Introduction

The first experiment demonstrated that a text-based agent can be effective in reducing user frustration. In the second experiment reported here, we wanted to investigate whether the effectiveness of an affective agent could be improved by changing the characteristics of the agent display. In Klein et al. (2002) and in our first experiment the agents were implemented as a text-only interactions. However, given that the agent seems to succeed through its use of a human–human communication strategy, it might be even better if the agent appeared more human-like. The second experiment, therefore, set out to investigate what effect using an embodied agent to provide the emotional feedback would have on the effectiveness of the affective agent.

We predicted that interaction with the embodied affective agent would lead to significant reductions in self-reported frustration levels and would be more effective at relieving frustration than a text-only agent. The hypotheses investigated in the experiment were:

- H1: participants interacting with the text-based affective agent will experience significant reductions in their self-rated frustration levels,
- H2: participants interacting with the embodied affective agent will experience significant reductions in their self-rated frustration levels,
- H3: participants interacting with the embodied affective agent will experience greater reductions in self-rated frustration levels than those interacting with the text-based agent.

4.2. Method

4.2.1. Procedure and experimental design

A between-groups experimental design was used. In one condition, participants interacted with a text-based affective agent with identical behaviour to that described in Study 1 above. In the second condition, participants interacted with an embodied affective agent. The outputs provided were the same as the text-based agent, but instead of appearing in a text-box, they appeared to come from an on-screen character via a speech bubble. The agent was a blond female programmed using Microsoft® Agent (http://www.davidware.com/freemsagents.html). Her appearance is shown in Fig. 3. This agent appearance was chosen on the basis of a small set of informal interviews in which a sample

Fig. 3. Agent character used in study 2.

of people were told about the application and then shown examples of different agents and asked to choose their favourite. In both conditions participants interacted with the same deliberately frustrating computer game that was used in Study 1.

The procedure and dependant variables were the same as Study 1.

4.2.2. Participants

Ten participants took part, with five assigned to each condition. In the affective agent condition there were 3 males and 2 females; in the text condition there were 3 females and 2 males. All participants were university students.

4.3. Results

The frustration rating results are shown below in Fig. 4.

In the text-based affective agent condition the mean frustration rating before the agent feedback (time 1) was 3.8; the mean frustration rating after the intervention (time 2) was 2.6. A Wilcoxon signed-ranks test was used to compare the frustration ratings at times 1 and 2. Since, we hypothesised that frustration would be lower following the agent's intervention, a one-tailed test was used. The result showed a significant difference between the ratings at times 1 and 2 ($W=0$, $n=5$, $P<0.05$). We were, therefore, able to accept H1 that participants interacting with the text-based affective agent will experience significant reductions in their self-rated frustration levels. This replicates the effect found in Study 1.

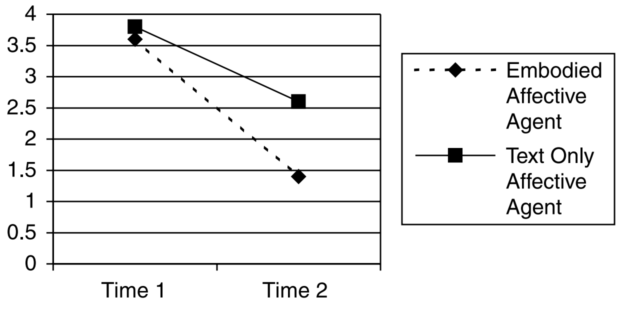

Fig. 4. Frustration reduction with embodied vs. text only agents in study 2.

In the embodied agent condition the mean rating of frustration at time 1 was 3.6 and at time 2 was 1.4. A one-tailed Wilcoxon signed-ranks test was used to compare the frustration ratings at times 1 and 2. The result showed a significant difference between the ratings at times 1 and 2 ($W=0$, $n=5$, $P<0.05$). We were, therefore, able to accept H2 that participants interacting with the embodied affective agent will experience significant reductions in their self-rated frustration levels.

In order to test H3 a one-tailed Man–Whitney test was used to compare the change in frustration in the text-based affective agent condition with the change in frustration in the embodied affective agent condition. The mean decrease in frustration in the text-based affective agent condition was 1.2. The mean decrease in frustration in the embodied agent condition was 2.2. The difference between the conditions was found to be significant ($U(5,5)=4$, $P<0.05$). We were, therefore, able to accept H3 that participants interacting with the embodied affective agent will experience greater reductions in self-rated frustration levels than those interacting with the text-based agent.

4.4. Conclusion

This experiment confirmed the frustration reducing effect of the text-based affective agent. It also showed that an embodied affective agent exhibiting the same active listening/empathic approach could reduce user frustration. Finally, it demonstrated that the embodied agent was more effective at reducing user frustration than the text-only version.

5. Study 3: agent gender

5.1. Introduction

The results of the second experiment suggest that affective feedback may be more effective at reducing frustration when delivered by a human-like character. We hypothesise that this is because there is a good match between the characteristics of the feedback strategy (human–human) and the characteristics of the entity delivering that feedback. Experiment 2 had used a female agent embodiment because this had been the preferred image selected in informal pre-trial interviews. For this final study, we wondered whether this selection of a female agent by our pre-trial participants might actually have had a meaningful effect on the results obtained. That is, might agent gender affect frustration reduction? Females are stereotypically considered more empathic than males. It could, therefore, be argued that frustration relieving feedback may appear more appropriate coming from a female character than a male character. The third experiment, therefore, investigates whether a female agent character will be more effective at relieving user frustration than a male agent character. The hypotheses investigated in this experiment were:

H1: participants interacting with the male embodied affective agent will experience significant reductions in their self-rated frustration levels,
H2: participants interacting with the female embodied affective agent will experience significant reductions in their self-rated frustration levels,
H3: participants interacting with the female embodied affective agent will experience greater reductions in self-rated frustration levels than those interacting with the male text-based agent.

5.2. Method

5.2.1. Procedure and experimental design

A between-groups experimental design was used. One group of participants interacted with a female agent and a second group of participants interacted with a male agent. The agent interactions were implemented in Microsoft Agent and Visual Basic. The behaviour of the agent was the same as in Studies 1 and 2 reported above. The embodiments chosen for this study were cartoon-type characters rather than the more realistic looking character used in study 2. A small pilot study (using 5 participants) was used to select the agent appearances from a selection of those freely available (see http://www.cantoche.com/english/gallery/msagent.htm for 'James', the male agent used and http://www.stegami.com/ for 'vrgirl', the female agent used). The chosen appearances are illustrated in Fig. 5. In both conditions participants interacted with the same frustrating game described in Study 1. The procedure was also the same as in Studies 1 and 2.

Fig. 5. Female and male embodiments used in study 3.

The dependent variable was frustration rating. This dependent variable was measured at two points during the experiment: time 1 was during the game, time 2 was after the game. A $2\times2\times2$ mixed factorial ANOVA was used to analyse the experimental data. The repeated measures factor was frustration rating time (time 1, time 2). The between subjects factors were gender of the agent and gender of the participant (included as a control variable). Change in frustration level (frustration at time 1 minus frustration at time 2) was used as the dependent variable for post hoc and planned comparisons of the between-subjects conditions, since this measure controls for differences in initial levels of frustration.

5.2.2. Participants

There were 10 participants in each condition. The groups were balanced for participant gender (with 5 males and 5 females per condition). All participants were university undergraduate students.

5.3. Results

ANOVA showed a significant main effect for frustration rating time ($F=144.1$, d$f=1$, $P<0.01$) indicating a reduction in frustration levels after interacting with the agent. Post hoc paired t-tests showed that frustration was reduced between the first rating and second rating for both agent types ($t=6.7$, d$f=9$, $P<0.01$ for the male agent; $t=8.82$, d$f=9$, $P<0.01$ for the female agent). This supports H1 and H2.

The ANOVA indicated a significant interaction effect between rating time and agent gender ($F=5.2$, d$f=1$, $P<0.05$), indicating that the male and female agents had differing effects on frustration reduction. A planned one-tailed t-test was used to test the hypothesis that female agents would be more effective at relieving frustration. This showed a significant difference ($t=-2.1$, d$f=18$, $P<0.05$) between the effectiveness of the male agent (mean frustration reduction 1.5) and female agent (mean frustration reduction 2.2). This result supports H3. A plot of the interaction is shown in Fig. 6.

The ANOVA also showed a significant interaction effect between rating time and participant gender ($F=5.2$, d$f=1$, $P<0.05$) indicating that the agents had differential effects on the changes in frustration observed for female and male participants. The interaction plot is shown in Fig. 7. This plot shows that at time 1, the female participants had, on average, slightly higher levels of frustration than the male participants. At time 2,

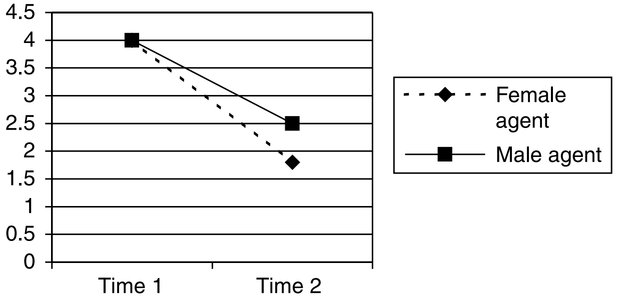

Fig. 6. Interaction plot showing the reduction in frustration with the male and female agent embodiments in study 3.

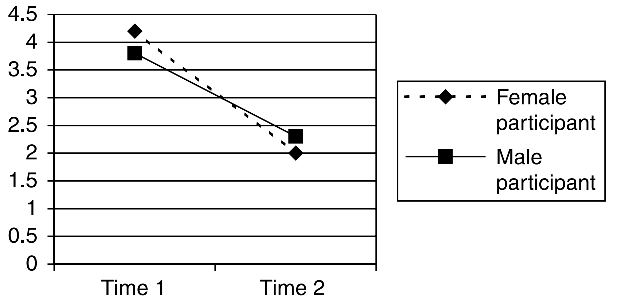

Fig. 7. Interaction plot showing the reduction in frustration for male and female experimental participants in study 3.

the situation is reversed, with the female participants showing slightly lower levels of frustration that the male participants. Overall this pattern suggests that the agent may have been more effective at lowering the frustration levels of female participants compared to male participants. A post hoc between-subjects comparison of frustration reduction for males and females was close to reaching statistical significance ($t=2.2$, $df=18$, $P=0.051$).

There was no significant three way interaction effect, showing that there is no evidence to suggest that male and female agents affect male and female users differently.

5.4. Conclusion

The experiment confirmed previous findings that affective agents can reduce user frustration. In addition, the results suggest that, as predicted, a female agent character is more effective than a male agent character at reducing frustration. An unexpected finding from the current study was that affective agents affected males and females differently. Although not statistically significant, there is preliminary evidence to suggest that affective agents may be more effective at relieving the frustration of female participants than at relieving the frustration of male participants.

6. Discussion

Collectively, the three studies reported here suggest that affective agents can be effective at reducing user frustration. Unlike previous work by Klein et al. (2002), we have been able to demonstrate an effect on users' subjective ratings of their own frustration level. We would suggest that our method may have been able to demonstrate this effect because frustration ratings were collected online during (and just after) task performance, rather than asking participants to provide post hoc judgements of their level of frustration level. The work reported here has also begun to explore how the characteristics of an affective agent can affect how good the agent is at reducing user frustration. Our results suggest that frustration reduction is improved if the agent is embodied, and that a female embodied agent may be more effective than a male agent character. These results were in line with the predictions that we had made on the basis of the improved match in these cases between the characteristics of the affect delivery mechanism (the agent) and the characteristics of the message. Thus, the feedback strategy used by the agent (affective feedback including active listening) is a technique derived from human–human interactions. Since, the technique is intrinsically human in quality it may appear more appropriate (and hence be more effective) coming from a human-type character. Similarly, the female character may be more effective than the male character because of the existing association between female gender and qualities such as empathy. Previous research suggests that gender stereotypes from the real world can transfer to human–computer interaction (Reeves and Nass, 1996; Lee et al., 2000) and this could well be happening here.

An unexpected outcome of the research reported here was that the agents appeared to affect female and male participants differently (i.e. there was a significant interaction effect between changes in frustration ratings over time and participant gender). There is tentative evidence to suggest that female participants may be affected more positively by the affective agents than male participants. This finding is in contrast to Klein et al. (2002) who checked for gender effects in their experiment but found none. It could be that the explanation for this difference in findings lies with the nature of the dependent variables used in our two studies. Thus, we used a subjective measure of frustration reduction while Klein et al. (2002) used a behavioural measure (time spent in continued interaction with the game). It could, therefore, be that the males in our study were just more reluctant to admit to extremes of emotion or to having been affected by the affective intervention compared to our female participants. This could be an instance of the well known self-serving bias in questionnaire usage where participants try to put themselves in a good light. In this case, men might wish to appear to conform to the male gender stereotype of being non-emotional. On the other hand our results could be indicative of a true gender difference, which the behavioural measure used in Klein et al. (2002) was not sensitive enough to demonstrate. Further work is needed to tease out these possibilities. In the mean time our results suggest that it is important to control for participant gender in experiments in this area.

Having concluded from Study 3 that participant gender may play a role in responses to affective agents, it is necessary to briefly revisit the design of Study 2. In this experiment, we had been unable to completely control for gender (recall that there were three males and two females in the embodied agent condition, and two males and three females in the text only condition). Luckily, the distribution of males and females within the two conditions means that the results are unlikely to have been confounded by gender effects. Thus, if, as we suspect, affective agents might be more effective for female participants, then we would expect (in the absence of experimental effects due to the agent types) the group with the largest proportion of females to experience the greater average drop in frustration levels. In fact, the opposite was true and it was the experimental group with the largest proportion of males that experienced the greater drop in frustration. This suggests that it is safe to attribute the main effects of the between groups comparison to the experimental manipulation of agent type.

While the results reported in this paper suggest a positive role of computer applications that try to relieve user frustration, there are a number of limitations to the work which must be acknowledged and discussed. The most obvious limitation, applying to the first and second studies reported here, relates to the small sample sizes used. These small sample sizes meant that multiple non-parametric statistical tests were needed in order to test the experimental hypotheses. Ideally, one global test (ANOVA) would be used to analyse all the data from each experiment (as was done in experiment 3). It must be recognised that by performing separate tests of the within and between subjects factors we slightly increase the possibility of a type I error occurring within the experiment as a whole (i.e. supporting an experimental hypothesis when in fact it is not true). On the other hand, the fact that we were able to get statistical support at all for all of our experimental hypotheses despite using such small samples suggests that these effects are fairly robust. In addition we only tested a priori hypotheses which to some extent mitigates against the risk of type I errors. Overall the cumulative support over several different studies for the experimental hypotheses concerning the ability of agents to reduce user frustration is highly encouraging.

A further limitation in this work, which applies to experiments 2 and 3, concerns the specific choices of agent embodiment used. For instance, in experiment 2 we cannot be sure that it was the humanness of the agent as such which had a positive effect, or whether the agent was effective because it had an attractive appearance (a blond female). Further work is needed to compare the roles of attractiveness and humanness in determining the effectiveness of such agents. In experiment 3, the two agents used differed according to more attributes than simply their gender. For example, their style of dress is different, they are drawn in different styles and their mannerisms vary. These differences are a consequence of our decision to use off-the-shelf agents to provide the experimental stimuli in our study. Future work needs to control for extraneous aspects of appearance such as this much more carefully before we can unambiguously attribute differences in effectiveness to agent gender alone. However, the fact that we have been able to establish that variations in agent appearance can affect how well the agent reduces frustration is an important finding in itself and suggests that further work in this area is warranted.

This paper has taken Klein et al. (2002) work as a starting point; that is it has looked at the basic premise that providing feedback tailored to observed affect can reduce user frustration. It has then explored how changing the source of the emotional feedback (agent appearance) affects the effectiveness of this approach. However, it should be recognised that a number of authors have begun to question whether the approach suggested by Klein et al. (2002) is the most appropriate way to deal with user emotion. Oatley (2004), for instance, has questioned whether simply expressing empathy for a user's frustration is enough; might an agent not be more effective still if it could also help to resolve the source of the user's frustration? This is clearly a question deserving of further research. Ward and Marsden (2004) query the assumption used in Klein et al. (2002) that the computer should be 'in charge' of the emotional interaction. Thus, Klein et al. (2002) assume that in the future emotion recognition technology will be used to detect user frustration and that this will trigger the deployment of an agent. Ward and Marsden (2004), on the other hand, argue that this one sided view does not represent how emotion works in the real world, where people control their emotional displays with deliberate communicative intent. They, therefore, argue for affective computing applications where users are given the freedom to use affect intentionally, taking the view that users must remain in control. Interestingly in the work reported here a key change from Klein et al. (2002) experimental methodology was that we allowed users to choose when to interact with the agent (rather than have the agent appear after a set period of interaction). This manipulation was intended to ensure that users were actually frustrated before they began an interaction with the agent. We had hoped this would better simulate the effect of future affective technology, which would not be triggered until frustration had actually been detected. However, another effect of our experimental manipulation was to hand control of registering frustration to the user (rather than the system). In light of Ward and Marsden's (2004) discussion it might be that this control contributed to the experimental effects that we were able to demonstrate. Future work could, therefore, usefully compare the effectiveness of agent interactions triggered by users themselves (through either the selection of the agent or through intentional affective acts) and those triggered by a recogniser picking up on underlying affect.

Acknowledgements

Thanks are due to Tony Elliman for creating the frustrating game used in these trials. Thanks also to the following students of Information Systems and Computing at Brunel University: Martin Saffu, Farah Aktar and Ashuveen Bhadal for their roles in running the three experiments reported in this paper.

References

Argyle, M., 1988. Bodily Communication, second ed. Methuen. New York.

Ark, W., Dryer, D.C., Lue, D.J., 1999. The emotion mouse. HCI International 99: Eighth International Conference on Human–Computer Interaction, Munich, Germany, August 22–27, 1999.

Belavkin, R.V., 2001. The role of emotion in problem solving. Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing, Heslington, NY, England, pp. 49–57.

Cohn, J.F., Kinade T., in press (available online at: http://www.psychology.pitt.edu/research/publications/cohn_AFA_20April2004.pdf). Use of automated facial image analysis for measurement of emotion expression. In: Coan, J.A., Allen, J.B. (Eds.), The Handbook for Emotion Elicitation and Assessment. Oxford University Press Series in Affective Science. OUP, New York.

Dehn, D.M., Van Mulder, S., 2000. The impact of animated interface agents: a review of empirical research. International Journal of Human–Computer Studies 52, 1–22.

Ekman, P., 1982. Emotion in the Human Face, second ed. Cambridge University Press, Cambridge, MA.

Ekman, P., Friesen, W.V., 1978. Facial Action Coding System. Consulting Psychologist Press, Palo Alto, CA.

Healey, J., Picard, R., Dabek, F., 1998. A new affect-perceiving interface and its application to personalized music selection. Proceedings of the 1998 Workshop on Perceptual User Interfaces, San Francisco, CA, November 4–6, 1998.

Klein, J., Moon, Y., Picard, R.W., 2002. This computer responds to user frustration: theory, design and results. Interacting with Computers 14 (2), 119–140.

Lee, E.-J., Nass, C., Brave, S., 2000. Can computer-generated speech have gender? An experimental test of gender stereotype. CHI'00 Extended Abstracts on Human Factors in Computing Systems, pp. 289–290.

Lisetti, C.L., Schiano, D.J., 2000. Automatic facial expression interpretation: where human–computer interaction, artificial intelligence and cognitive science intersect. Pragmatics and Cognition (Special Issue on Facial Information Processing) 8 (1), 185–235.

Nakatsu, R., Nicholson, J., Tosa, N., 1999. Emotion recognition and its application to computer agents with spontaneous interactive capabilities. Proceedings of the Seventh ACM International Conference on Multimedia, Orlando, Fl, USA, October 30–November 5, 1999, pp. 343–351.

Oatley, K., 2004. The bug in the salad: the uses of emotions in computer interfaces. Interacting with Computers 16 (4), 693–696.

Picard, R., 1997. Affective Computing. The MIT Press, Cambridge, MA.

Picard, R.W., 1998. Toward agents that recognize emotion. Actes Proceedings IMAGINA, Monaco, March 1998, pp. 153–165.

Reeves, B., Nass, D., 1996. The Media Equation: How People Treat Computers, Television and New Media like Real People and Places. CSLI Publications, Stanford.

Scheirer, J., Fernadez, R., Picard, R.W., 1999. Expression glasses: a wearable device for facial expression recognition. Proceedings CHI, Late-breaking Paper, Pittsburgh, PA, February, 1999.

Schiano, D.J., Ehrlich, S.M., Rahardja, K., Sheridan, K., 2000. Face to InterFace: facial affect in (hu)man and machine. Proceedings of CHI 2000, pp. 193–200.

Ward, R.D., Marsden, P.H., 2004. Affective computing: problems, reactions and intentions. Interacting with Computers 16 (4), 707–713.