Cogn Tech Work (2004) 6: 4–14
DOI 10.1007/s10111-003-0143-x

ORIGINAL ARTICLE

Fatma Nasoz · Kaye Alvarez · Christine L. Lisetti
Neal Finkelstein

Emotion recognition from physiological signals using wireless sensors for presence technologies

Received: 16 June 2003 / Accepted: 15 October 2003 / Published online: 19 December 2003

Abstract In this article we describe a new approach to enhance presence technologies. First, we discuss the strong relationship between cognitive processes and emotions and how human physiology is uniquely affected when experiencing each emotion. Secondly, we introduce our prototype multimodal affective user interface. In the remainder of the paper we describe the emotion elicitation experiment we designed and conducted and the algorithms we implemented to analyse the physiological signals associated with emotions. These algorithms can then be used to recognise the affective states of users from physiological data collected via non-invasive technologies. The affective intelligent user interfaces we plan to create will adapt to user affect dynamically in the current context, thus providing enhanced social presence.

Keywords Emotion recognition · Social presence · User interfaces

1 Introduction and motivation

The main thrust of this research is developing computer systems that can recognise its users' emotional states and adapt to them accordingly in order to enhance social presence. One of the first authors on presence was Marvin Minsky (1980), who coined the term *telepresence,* referring to the human operator's sense of being at a remote real environment in a teleoperation. In more recent work on presence, the environments in consideration are both virtual as well as real. Ijsselsteijn et al. (2000) define presence as the user's perception of "being there" in a mediated environment, while Lombard and Ditton's (1997) definition of *presence* is "the perceptual illusion of nonmediation". *Co-presence* is defined by Casanueva and Blake (2001) as the user's sense that (1) there are other participants existing in the virtual environment, and (2) s/he is having interaction with real people.

Many research studies on presence show that users are responding socially and emotionally to the systems, the characters in the virtual environments (VEs), or the robots that they are interacting with. For example, emotions affect the user's perception of the VE (Ijsselsteijn 2002; Lombard and Ditton 1997), and the interaction with the VE, in turn, affects the user's emotions. (Dillon et al. 2000; Kalawsky 2000).

In addition to the presence of emotions in presence technologies, emotions are essential for human thought processes that influence interactions between people and intelligent systems. Different aspects of cognition affected by emotions include: perception and organisation of memory (Bower 1981); categorisation and preference (Zajonc 1984); goal generation, evaluation, and decision-making (Damasio 1994); strategic planning (Ledoux 1992); focus and attention (Derryberry and Tucker 1992); motivation and performance (Colquitt et al. 2000); intention (Frijda 1986); communication (Birdwhistle 1970; Ekman and Friesen 1975; Chovil 1991); and learning (Goleman 1995).

The strong interface between emotion and cognition and the effects of emotion on humans' performances in VEs make it desirable, if not necessary, to create intelligent computer systems that understand users' emotional states, learn their preferences, and respond accordingly. There are various applications for such systems, including training, driving safety, and telemedicine.

F. Nasoz (✉) · C. L. Lisetti
Department of Computer Science,
University of Central Florida,
Orlando, FL 32816-2362, USA
E-mail: fatma@cs.ucf.edu
Tel.: 1-407-8233931

K. Alvarez
Personnel Board of Jefferson County,
Birmingham, AL, USA

N. Finkelstein
Simulation Technology Center,
12423 Research Parkway,
Orlando, FL 32826-3276, USA

1.1 Training/learning

Learning is a cognitive process that can be affected by one's emotional state. For example, *frustration* can lead to negative attitudes towards the training stimulus (Rozell and Gardner 2000) and can reduce a person's belief in his or her ability to do well in training (Briggs et al. 1998). As a result, frustration can hamper learning (Lewis and Williams 1989). Learning can also be impaired when trainees are experiencing high levels of *anxiety* during training. In training situations, anxiety is presumed to interfere with the ability to focus cognitive attention on the task at hand because that attention is preoccupied with thoughts of past negative experiences with similar tasks, in similar situations (Martocchio 1994; Warr and Bunce 1995).

With the affective intelligent user interfaces we are creating, we aim to enhance *presence* and *co-presence* in learning environments by teaching the system to recognise the user's affective state and adapt its processes in order to aid their learning. For example once the system learns a user's preferences and emotional states, when the user is in a learning environment and becomes anxious, in response, the system can provide the user's preferred style of encouragement, thus potentially reducing anxiety and allowing the learner to focus more attention on the task. Similarly, when the system recognises that the learner is becoming frustrated or bored, it could adjust the pace of the training accordingly so that the optimal level of arousal for that user's learning is achieved. In this manner, the system will provide assistance to the learner in order to enhance positive attitudes and emotions, therefore enhancing their learning (Lorenz et al. 2000; Martocchio 1994; Martocchio and Dulebohn 1994; Martocchio and Judge 1997). All these adaptation techniques will improve the learner's sense of being in a real classroom environment where a live instructor would typically recognise these same emotions and respond accordingly, thus enhancing presence.

1.2 Telemedicine

Tele-home health care (tele-HCC) has been practiced in the United States for over a decade. Tele-HHC provides communication between medical professionals and patients in cases where hands-on care is not required, but regular monitoring is necessary. For example, tele-HHC interventions are currently used to collect vital sign data remotely (e.g., ECG, blood pressure, oxygen saturation, heart rates, and respiration), verify compliance with medicine and/or diet regimes, and assess aspects of mental or emotional status (Allen et al. 1996; Crist et al. 1996; Darkins and Carey 2000; Warner 1997). However, formulating an assessment can be particularly difficult in tele-HHC settings where patients are treated and monitored using multiple media devices that filter out important social and emotional cues (e.g., facial expressions).

With the affective intelligent user interfaces we are creating, we aim to enhance *telepresence* and *presence* in telemedicine environments. For example, when communicating in a Tele-HHC environment, our system's avatar could mimic the facial expressions of users at both sites (Lisetti et al. 2003). During this interaction, when the system recognises and then transmits data indicating the patient is experiencing depression or sadness, health-care providers monitoring them will be better equipped to respond. Such a system has the potential to improve patient satisfaction and health. That is, not only emotional information can provide key information regarding a patient's mental or physical health and increase patient satisfaction as a result of more empathic communications, but the power of positive emotions themselves have shown to be beneficial during the recovery process (Damasio 1994). Social presence during patient-physician communication is indeed essential; furthermore, the rising use of Tele-HHC signifies a need for efforts aimed at enhancing such presence.

1.3 Driving safety

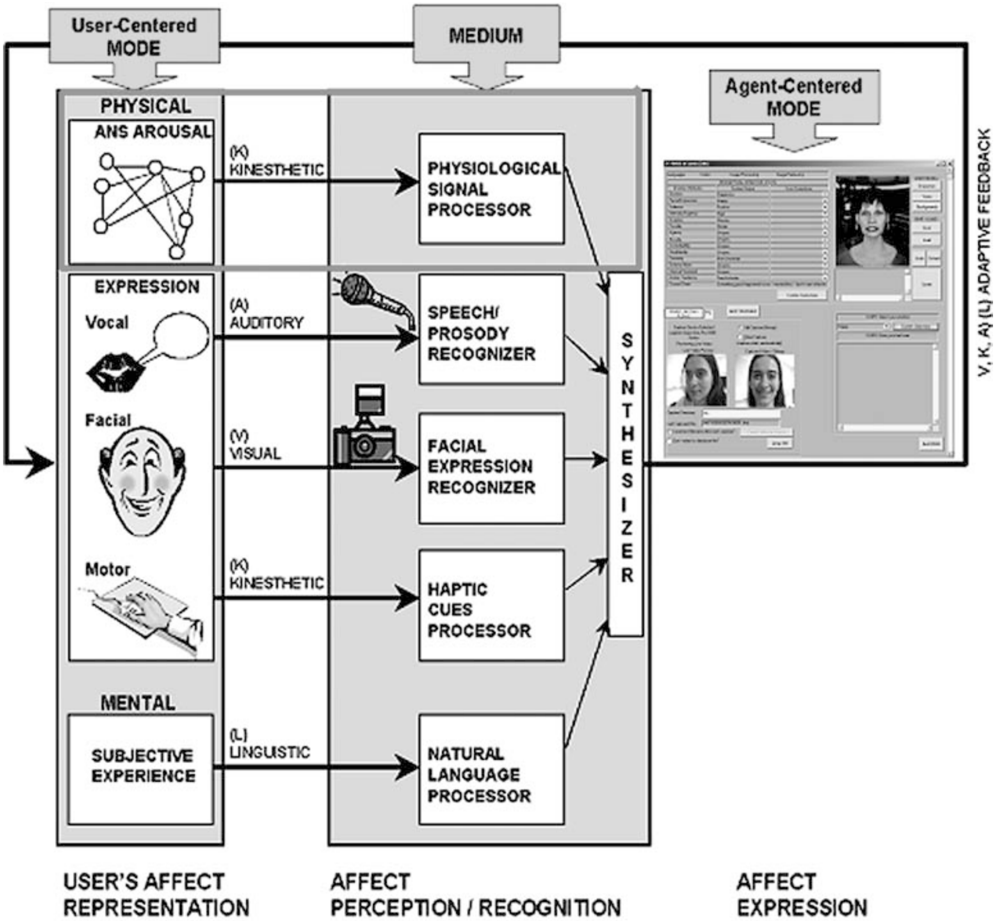
The inability to manage one's emotions while driving is identified as one of the major causes for accidents (James 2000). When drivers become angry, their thinking, perception, and judgments are impaired, thus leading to the misinterpretation of events. In addition, drivers often lack the ability to calm themselves when they are angry or frustrated. With the affective intelligent user interfaces we are creating, we aim to enhance *co-presence* in the driving environment. For example, when the system recognises the driver in a state of frustration, anger, or rage, it could change the music (James 2000), or suggest a relaxation technique (Larson and Rodriguez 1999), depending on the driver's preference. The assistance provided by such a system would enhance driver's perception of social presence.

These three applications suggest that by enhancing the social and emotional cues in human-computer interactions, users may benefit from improved satisfaction in learning and healthcare, increased skills in e-learning environments, and assistance with safe driving habits. Furthermore, technological interactions that allow for social presence may increase human acceptance of such systems beyond that of typical cold technologies.

2 Increasing social presence with the multimodal affective user interface

The first step in our approach for enhancing social presence focuses on accurately recognising the user's emotional state. Fig. 1, which was originally proposed by Lisetti (1999) , shows the framework of our approach (Bianchi and Lisetti 2002, Lisetti and Nasoz 2002). This architecture is designed to: (1) have a database of

Fig. 1 Human multimodal affect expression matched with multimedia computer sensing, from Lisetti (1999)



emotion concepts for each emotion expressed by a given user; (2) gather multi-modal physiological signals and expressions of a specific emotion to recognise the user's emotional state; (3) provide feedback to the user about his or her emotional state; and (4) dynamically adapt its interface to the user's current affective state and do so according to the user's preferences of interaction in similar contexts or applications.

In order to make emotion recognition accurate and reliable, our completed system will take as input both physiological components (facial expressions, vocal intonation, skin temperature, galvanic skin response (GSR), and heart rate) and subjective components (written or spoken language) that are associated with emotions experienced by the user.

Currently, we are working on recognising users' emotions with non-invasive technologies measuring physiological signals of autonomic nervous system arousal (skin temperature, heart rate, and GSR), which are then mapped to their corresponding emotions (represented by the red rectangle in Fig. 1).

Based on this architecture, we designed the prototype MAUI: multimodal affective user interface (Fig. 2) (Lisetti and Nasoz 2002) which is used as our in-house research tool. In the following sections we will briefly describe the main features of MAUI. As we mentioned, MAUI is currently our research tool and it will be adapted to the various needs of each application (training/learning, etc.) before it is introduced to be the final user interface.

2.1 Avatar

The upper right section of MAUI displays an anthropomorphic avatar that adapts its facial expressions and vocal intonation according to the user's affective state. Earlier studies have emphasised that facial expressions are universally expressed and recognised by humans (Ekman 1989). In addition, the human face is considered an independent channel of communication that helps to co-ordinate conversations in human-human interactions (Takeuchi and Nagao 1993). In human-computer interactions, research suggests that having an avatar as part of an interface helps to increase human performance. For example, Baylor's study (2000) emphasised the increase in "motivational qualities" of learners by achieving human resemblance in agents of a learning environment.

Including expressive avatars also has the potential to increase perception of presence. For example, Casanueva and Blake (2001) found that the level of subjective co-presence was higher for subjects interacting with characters displaying gestures and facial expressions

Fig. 2 MAUI–multimodal affective user interface

than it was for subjects interacting with avatars communicating with a static neutral expression and no gestures.

The avatar in our MAUI system was created using Haptek PeoplePutty software (Haptek Inc.) and will have the ability to make context relevant facial expressions. The various ways the avatar can be used are: mirroring users' emotions as a method to confirm their emotional states (see Fig. 3); responding with socially appropriate facial expressions as users display their emotional states (i.e., avatar displays empathy when the user is frustrated by a task); assisting users in understanding their own emotional states by prompting them with simple questions and comparing the various components of the states they believe they are experiencing, with the system's output; and displaying the facial expressions of individuals in a text-based chat session in order to enhance communication. In addition, in order to address individual differences in user preferences, MAUI system provides a choice of avatar displays including different ages, genders, ethnic backgrounds, skin colors, voices, hair, make-up, accessories, and backgrounds.

2.2 User text input

The text field in the lower right hand corner of Fig. 2 is for users to communicate with the system via text. Users enter here the information about their own emotional

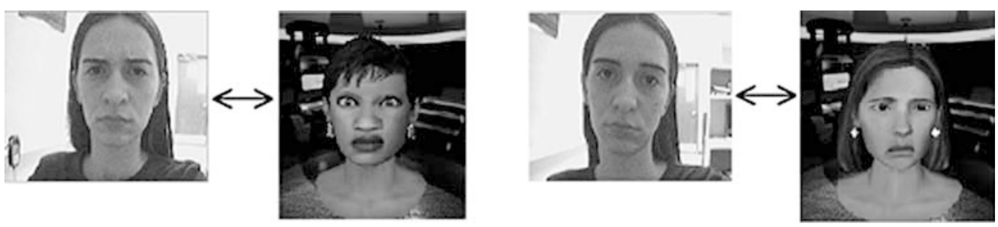


Fig. 3 Avatar mirroring user's anger and sadness respectively

states in their own words. Users' input will be interpreted by natural language understanding algorithms for more accurate recognition of emotions. Emotion recognition from natural language is already being implemented in intelligent computer systems (Guinn and Hubal 2003).

2.3 Ongoing video and captured image

The first image in the lower left hand corner of Fig. 2 displays the ongoing video of the user, which is recorded by a camera connected to user's computer. This video captured during interaction is saved in order to compare the system's interpretation of changes in the user's physiological arousal and the changes in her/his facial expressions over time. The second image displays the still image of the user captured at specific times for facial expression recognition.

2.4 Feedback to user

In the upper left hand corner of Fig. 2, the system displays, in a text format, its interpretation of the user's current emotional state (i.e., happy, sad, frustrated, angry, etc.) by indicating the emotion components (i.e., valence, intensity, causal chain, etc.) associated with the emotion, facial expression reading, and physiological signals. As mentioned previously, the information about the user's affective state that feeds the system is gathered by physiological measurements of arousal. These data are then interpreted through pattern recognition algorithms, which identify the user's current emotion.

3 Previous research on emotion recognition

Research conducted on understanding the connection between emotions and physiological arousal is growing. Manual analyses have been successfully used for this purpose (Ekman et al. 1983; Gross and Levenson 1997). However, interpreting the data with statistical methods and algorithms is beneficial in terms of actually being able to map them to specific emotions.

Collet et al. (1997) showed neutral and emotionally loaded pictures to participants in order to elicit happiness, surprise, anger, fear, sadness, and disgust. The physiological signals measured were: Skin conductance (SC), skin potential (SP), skin resistance (SR), skin blood flow (SBF), skin temperature (ST), and instantaneous respiratory frequency (IRF). Statistical comparison of data signals was performed pair-wise, where 6 emotions formed 15 pairs. Out of these 15 emotion pairs, electrodermal responses (SR, SC, and SP) distinguished 13 pairs, and similarly combination of thermo-circulatory variables (SBF and ST) and Respiration could distinguish 14 emotion pairs successfully.

Picard et al. (2001) used personal imagery technique to elicit happiness, sadness, anger, fear, disgust, surprise, neutrality, platonic love, and romantic love. The physiological signals measured were GSR, heartbeat, respiration, and electrocardiogram. The algorithms used to analyse the data were sequential forward floating selection (SFFS), fisher projection, and a hybrid of these two. The best classification achievement was gained by the hybrid method, which resulted in 81% overall accuracy.

Table 1 summarises results of studies investigating the relationship between emotions and physiological arousal by using other statistical procedures such as ANOVA and hidden Markov models. All these studies succeeded in finding a pattern of physiological signals for each of the emotions elicited. In summary, the results of these studies suggest that physiological patterns can successfully be identified using statistical procedures.

4 Our emotion elicitation and recognition study

In the following sections, we describe a pilot study and a subsequent emotion-physiological investigation with the goal of capturing the physiological signals associated with particular emotions. To elicit the target emotions in our study, we used short segments of movies and short films. In their study, Gross and Levenson (1995) reported the most effective films found to elicit discrete emotions. Films selected for investigation were subjected to the following criteria: (1) the length of the scene needed to be relatively short, (2) the scene needed to be understood without explanation, and (3) the scene needed to elicit a single emotion. Out of 78 films, the two movie scenes resulting in the highest subject agreement for eliciting discrete emotions were presented. Hit rates for these films were: amusement (84% and 93%), disgust (85% and 80%), sadness (94% and 76%), surprise (75% and 67%), fear (71% and 60%), contentment (58% and 43%), and anger (22% and 42%). As can be seen, the amusement, disgust, and sadness scenes were most successful in producing the target emotion. The more difficult emotions to elicit were anger, contentment, and fear. Upon further analyses, the authors reported that contentment films elicited high degrees of happiness; anger films were affiliated with a host of other emotions, including disgust; and fear films were confounded with tension and interest. However, the authors concluded that, if the goal is to elicit one emotion more intensely than others, films are a viable choice. These findings also suggest that, in natural environments, emotions most likely do not occur in isolated episodes. Therefore, in order to increase social presence, we believe that detection of emotional cues from physiological data must also be gathered in a natural environment rather than one where emotions are artificially extracted from other naturally co-occurring states. The following sections describe our study in more detail.

Table 1 Previous research on recognising emotion from physiological signals

Author	Emotion elicitation method	Emotions elicited	N	Measures	Data analyse technique	Results
Lanzetta JT, Orr SP (1986)	Vocal tone, slide of facial expressions, electric shock	Happiness and fear	60 (23 female, 37 male)	Skin conductance (galvanic skin response)	ANOVA	Fear produced a higher level of tonic arousal and larger phasic skin conductance.
Vrana SC, Cuthbert BN, Lang PJ (1986)	Imagining and silently repeating fearful and neutral sentences	Neutral and fear	64	Heart rate and self report	ANOVA and Newman-Keuls pairwise comparison	Heart rate acceleration was higher during fear imagery than neutral imagery or silent repetition of neutral sentences or fearful sentences.
Pecchinenda A, Smith C (1996)	Difficult problem solving	Difficult problem solving	32 (16 male, 16 female)	Skin conductance, self-report, and objective task performance	ANOVA, MANOVA, and correlation/regression analyses	Within trials, skin conductance increased at the beginning of the trial, but decreased by the end of the trials for the most difficult condition.
Sinha R, Parsons O (1996)	Imagery script development	Neutral, fear, joy, action, sadness, and anger.	27 males (ages 21–35)	Heart rate, skin conductance, finger temperature, blood pressure, electro-oculogram, and facial electromyograms	Discriminant function analyses and ANOVA	%99 correct classification was obtained. This indicates that emotion-specific response pattern for fear and anger is accurately differentiable from each other and from neutral.
Scheirer J, Fernandez R, Klein J, Picard RW (2002)	A slow computer game interface	Frustration	36	Skin conductivity and blood volume pressure	Hidden Markov models	Pattern recognition worked significantly better than random guessing while discriminating between regimes of likely frustration from regimes of much less likely frustration.

4.1 Pilot panel study

Before collecting physiological data, we conducted a pilot panel study with movie scenes resulting in high subject agreement from Gross and Levenson's (1995) work. Because some of their movies were not obtainable and anger and fear movie scenes evidenced low subject agreement, alternative clips were also investigated. The purpose of the panel study was to determine the movies that may result in high subject agreement for our subsequent study, which will be described shortly. The following sections describe the panel study and results.

Sample. The sample included 14 undergraduate and graduate students from the psychology and computer science departments of a university in Florida. There were 7 females and 7 males: 10 Caucasians, 1 Hispanic American, 1 African American, and 2 Asians. Their ages ranged from 18 to 35. Specific ages were not requested; therefore a mean age was not calculated.

Movie clips. Emotions were elicited using scenes from 21 movies. Seven movies were included in the analysis based on the findings of Gross and Levenson (1995; see Table 2). An additional 14 movie clips were found by the authors. The final sample included 4 movies targeted to elicit anger (*Eye for an Eye, Schindler*'s List, American History, and *My Bodyguard),* 3 movies to elicit sadness (*Powder, Bambi,* and *The Champ*), 4 to elicit amusement (*Beverly Hillbillies, When Harry Met Sally, Drop Dead Fred,* and *The Great Dictator*), 1 to elicit disgust (*Fear Factor*), 5 to elicit fear (*Jeepers Creepers, Speed, The Shining, Hannibal,* and *Silence of the Lambs*), and 4 to elicit surprise (*Jurassic Park, The Hitcher, Capricorn One,* and a homemade clip called *Grandma*).

Procedure. The 14 subjects participated as a group simultaneously. Once consent forms were completed, the participants were given questionnaires and asked to answer the demographic items before beginning the study. Then the subjects were informed that they would be watching scenes from various movies geared to elicit emotions. They were also told that between each movie they would be prompted to answer questions about the emotions they experienced as a result of the scene. Lastly, they were asked to respond according to the emotions *they* experienced and not the emotions displayed by the actors. A computerised slide show played the scenes and, after each of the 21 clips, a slide was presented asking the participants to answer the survey items for the prior scene.

Measures. The questionnaire included three demographic questions: age ranges (18–25, 26–35, 36–45, 46–55, or 56+), gender, and ethnicity. For each scene, 4 questions were asked. The first question asked what emotion they experienced from the video clip they viewed, and provided 7 options (anger, frustration, amusement, fear, surprise, sadness, and other). If the participant checked "other" they were asked to specify which emotion they experienced. The second question asked the participants to rate the intensity of the emotion on a 6 point scale. The third question asked participants if they had experienced any other emotions at the same intensity or higher, and if so, to specify what that emotion was. The final question asked participants if they had seen the movie in the past.

Results. The goal of the pilot study was to find the movie scenes that resulted in (a) 90% agreement or higher on the target emotion, and (b) 3.5 or higher average intensity. Table 3 lists the hit rates and average intensities for the clips with >90% agreement. There was not a movie with a high level of agreement for anger. With intensity in mind, Gross and Levenson's (1995) clips were most successful at eliciting the emotions in our investigation, except for anger. In their study, the movie with the highest hit rate for anger was *My Bodyguard* (42%). In our pilot study, the hit rate was 29% with a higher hit rate for frustration (36%). However, because anger is an emotion of interest for future research with driving simulators, we included the movie with the highest hit rate, *Schindler*'s List (hit rate was 36%, average intensity was 5.00). In addition, for amusement, the movie *Drop Dead Fred* was chosen to replace *When Harry Met Sally* due to the embarrassment experienced by some of the subjects when viewing the latter. The final set of movie scenes chosen for the study are presented in Table 4.

Table 2 Movies from Gross & Levenson (1995)

Emotion	Movie	N	Agreement	Mean intensity*
Sadness	Bambi	72	76%	5.35
	The Champ	52	94%	5.71
Amusement	When Harry Met Sally	72	93%	5.54
Fear	The Shining	59	71%	4.08
	Silence of the Lambs	72	60%	4.24
Anger	My Bodyguard	72	42%	5.22
Surprise	Capricorn One	63	75%	5.05

*8-point scale

Table 3 Hit rates and average intensities for movies with >90% agreement

Emotion	Movie	Agreement	Mean intensity	SD
Sadness	Powder	93%	3.46	1.03
	Bambi	100%	4.00	1.66
	The Champ	100%	4.36	1.60
Amusement	Beverly Hillbillies	93%	2.69	1.13
	When Harry Met Sally	100%	5.00	0.96
	Drop Dead Fred	100%	4.00	1.21
	Great Dictator	100%	3.07	1.14
Fear	The Shining	93%	3.62	0.96
Surprise	Capricorn One	100%	4.79	1.25
(N = 14)				

Table 4 Movie scenes selected for the study

Emotion	Movie	Scene
Sadness	The Champ	death of the Champ
Anger	Schindler's List	woman engineer being shot
Amusement	Drop Dead Fred	restaurant scene
Fear	The Shining	boy playing in hallway
Surprise	Capricorn One	agents burst through the door

4.2 Emotional signals data generation and collection

Sample. The final sample included 29 undergraduate students enrolled in a computer science course. There were 3 females and 26 males: 21 Caucasians, 1 African American, 1 Asian American and 6 individuals who did not report their ethnicity. Their ages ranged from 18 to 40 (19 individuals were 18–25 and 10 were 26–40). Specific ages were not requested, therefore a mean age was not calculated.

Procedure. One to three subjects participated in the study during each session. After signing consent forms, a non-invasive SenseWear armband (see Fig. 4) was placed on each subject's right arm to collect GSR, heart rate, and temperature for emotion recognition purposes. As the participants waited for the armband to detect their physiological signals, they were asked to complete a pre-study questionnaire. Once the armband signalled it was ready, the subjects were instructed on how to place the chest strap. After the chest straps were activated, the in-study questionnaire was placed face down in front of the subjects and the participants were told the following: (1) to find a comfortable sitting position and try not to move around until answering a questionnaire item, (2) that the slide show would instruct them to answer specific items on the questionnaire, (3) to please not look ahead at the questions, and (4) that someone would sit behind them at the beginning of the study to activate the armband.

A 45-minute computerised slide show was then activated. The study began with a slide asking the subjects to relax, breathe through their nose, and listen to soothing music. Slides of natural scenes were presented, including pictures of the oceans, mountains, trees, sunsets, and butterflies. These slides were presented for 6 seconds each. After 2.5 min, the first movie clip played (sadness). Once the clip was over, the next slide asked the participants to answer the questions relevant to the scene they watched. This slide stayed on screen for 45 seconds. Starting again with the slide asking the subjects to relax while listening to soothing music, this process continued for the anger, fear, surprise, frustration, and amusement clips. The frustration segment of the slide show asked the participants to solve analytical math problems without paper and pencil. The movie scenes and frustration exercise lasted from 70 to 231 seconds each. After the slide show ended, the participants were asked to remove their chest straps first and then the armbands. The in-study questionnaires were collected and the subjects were asked if they had any questions or comments.

Fig. 4 BodyMedia SenseWear armband

Measures. The pre-questionnaire included three demographic questions: age ranges (18–25, 26–35, 36–45, 46–55, or 56 +), gender, and ethnicity. The in-study questionnaire included 3 questions for each emotion. The first question asked if they experienced sadness (or the relevant emotions), and required a yes or no response. The second question asked the participants to rate the intensity of the emotion they experienced on a 6 point scale. The third question asked participants if they had experienced any other emotions at the same intensity or higher, and if so, to specify what that emotion was.

Self-report. Table 5 reports subject agreement and average intensities for each movie scene and the math problems. A two sample binomial test of equal proportions was conducted to determine whether the agreement rates for the panel study differed from the results obtained with this sample. Participants in the panel study agreed significantly more to the target emotion for the sadness and fear films. On the other hand, the subjects in this sample agreed more for the anger film. This lack of reliability in subject agreement across studies may be due to small sample sizes; however, other possible explanations will be provided in "Discussion and Future work" section.

4.3 Emotion recognition with machine learning

After determining the time slots corresponding to the point in the film where the intended emotion was most

Table 5 Hit rates and average intensities

Emotion	Movie	N	Agreement	Mean intensity	SD
Sadness	The Champ	27	56%	3.53	1.06
Anger	Schindler's List	24	75%	3.94	1.30
Fear	The Shining	23	65%	3.58	1.61
Surprise	Capricorn One	21	90%	2.73	1.28
Frustration	Math Problems	22	73%	3.69	1.35
Amusement	Drop Dead Fred	23	100%	4.26	1.10

likely to be experienced, the procedures described above resulted in the following set of physiological records: 24 for anger, 23 for fear, 27 for sadness, 23 for amusement, 22 for frustration, and 21 for surprise. The number of data sets for each emotion is different from the total sample size because for some participants, collected physiological signals data were not complete for every emotion. We stored the data in a three dimensional array of real numbers. The three dimensions are (1) the subjects who participated in the experiment, (2) the emotion classes (sadness, anger, surprise, fear, frustration, and amusement), and (3) the data signal types (GSR, temperature, and heart rate).

Each slot of the array consists of the normalised average value of one specific data signal belonging to one specific participant while s/he was experiencing one specific emotion. (e.g., a slot contains the normalised average skin temperature value of participant #1 while s/he was experiencing anger). We normalised the data for each emotion in order to calculate how much the physiological responses change as the participants go from a relaxed state to the state of experiencing a particular emotion. Normalisation is also important for minimising the individual differences of participants in terms of the physiological responses they give while experiencing a specific emotion.

The values of each data type were normalised by using the average value of corresponding data collected during the relaxation period for the same participant. For example, Eq. 1 shows how we normalised the GSR values:

$$normalized\_value\_GSR = \frac{original\_value\_GSR - relaxatio\_value\_GSR}{relaxation\_value\_GSR} \quad (1)$$

After storing the normalised data in the three dimensional array, we implemented three algorithms to analyse it: (1) k-Nearest Neighbor (KNN) (Mitchell 1997), (2) Discriminant Function Analysis (DFA) (Nicol 1999), and (3) Marquardt Backpropagation (MBP) (Hagan and Menhaj 1994).

As shown in Table 6, with KNN algorithm the recognition accuracy obtained was: 67% for sadness, 67% for anger, 67% for surprise, 87% for fear, 72% for frustration, and finally 70% for amusement.

As can be seen in Table 7, the results of the DFA algorithm demonstrated a similar pattern of accuracy across emotions to that of the KNN algorithm. The DFA algorithm successfully recognised sadness (78%), anger (72%), surprise (71%), fear (83%), frustration (68%), and amusement (74%).

As shown in Table 8, the recognition accuracy gained with MBP algorithm was: 92% for sadness, 88% for anger, 70% for surprise, 87% for fear, 82% for frustration, and 83% for amusement.

Table 6 Emotion recognition results with KNN algorithm

	Elicited emotion						
		Sadness	Anger	Surprise	Fear	Frustration	Amusement
Recognised emotion	Sadness	67%	8%	0%	0%	0%	0%
	Anger	5%	67%	0%	0%	4%	0%
	Surprise	7%	4%	67%	13%	4%	4%
	Fear	7%	8%	15%	87%	20%	13%
	Frustration	7%	13%	9%	0%	72%	13%
	Amusement	7%	0%	9%	0%	0%	70%

Table 7 Emotion recognition results with DFA

	Elicited emotion						
Recognised emotion		Sadness	Anger	Surprise	Fear	Frustration	Amusement
	Sadness	78%	8%	0%	4%	9%	0%
	Anger	4%	72%	5%	0%	0%	4%
	Surprise	4%	4%	71%	9%	5%	4%
	Fear	7%	8%	14%	83%	13%	9%
	Frustration	0%	4%	10%	4%	68%	9%
	Amusement	7%	4%	0%	0%	5%	74%

Table 8 Emotion recognition results with MBG algorithm

	Elicited emotion						
Recognised emotion		Sadness	Anger	Surprise	Fear	Frustration	Amusement
	Sadness	92%	0%	0%	0%	4%	0%
	Anger	8%	88%	9%	5%	10%	4%
	Surprise	0%	0%	70%	8%	0%	4%
	Fear	0%	8%	4%	87%	0%	0%
	Frustration	0%	4%	13%	0%	82%	9%
	Amusement	0%	0%	4%	0%	4%	83%

Overall, the DFA algorithm was better than the KNN algorithm for sadness, anger, surprise, and amusement. On the other hand, KNN performed better for frustration and fear. MBP algorithm performed better than both DFA and KNN for all emotion classes, except surprise.

4.4 Discussion and future work

Feldman Barrett et al. (2001) found that individuals vary in their ability to identify the specific emotions they experience (emotion differentiation). For example, some individuals clump all negative emotions together and all positive emotions together and are thus able to indicate whether the experience is unpleasant (negative emotion) or pleasant (positive emotion), but less able to report the specific emotion. Other individuals are able to discriminate between both negative (i.e., fear versus anger) and positive (happiness versus joy) emotions and are thus able to identify the specific emotion experienced.

The results we obtained by interpreting the physiological data with three different algorithms supported the above findings. For example, our algorithms recognised sadness with 67% (KNN), 78% (DFA), and 92% (MBP) accuracy rates, although only 56% of the participants reported that they experienced sadness. Similar results were obtained for surprise with KNN, DFA, and MBP; and for anger and frustration with MBP.

We plan to continue:

1. Conducting more experiments in different environments, such as virtual training environments and car simulators for various applications including soldiers' training, driving safety, and telemedicine
2. Measuring physiological signals associated with specific emotions using different equipment and generating comparison in terms of accuracies obtained by these equipment
3. Developing and fine-tuning pattern recognition algorithms for emotion recognition
4. Integrating different emotion recognition systems for various modalities such as facial expressions, vocal intonation, and natural emotion language understanding
5. Creating interaction models for the interface especially the avatar to adapt to the user
6. Building models of the emotional patterns of users for a more personalised adaptation of the system

5 Conclusion

We based our research on improving social presence of users in various virtual environments by recognising their emotions and interacting with them accordingly. As we discussed, there are different modalities of input from the user that can be used for automatic emotion recognition including physiological arousal, facial expressions, vocal intonation, and natural language. In this article, we described our research on emotion recognition by interpreting physiological signals. We conducted an experiment to elicit emotions and measure physiological data and we implemented three pattern recognition algorithms for the analyses of these data. We obtained very promising results when we interpreted our data with these algorithms. Overall, the KNN algorithm classified emotions by 71%, DFA by 74%, and MBP by 83% accuracy. Because emotions play an important role during interactions in VEs, the systems that recognise the emotional states of their users accurately and interact accordingly will help us develop environments with enhanced social presence.

References

Allen, A. Roman, L. Cox, R. and Cardwell, B. (1996) Home health visits using a cable television network: user satisfaction. J Telemed Telecare 2:92–94

Baylor AL (2000) Beyond butlers: Intelligent agents as mentors. J Educ Comput Res 22:373–382

Bianchi N, Lisetti CL (2002) Modeling multimodal expression of user's affective subjective experience. Int J User Model User-adapt Interact 12(1):49–84

Birdwhistle R (1970) Kinesics and context: essays on body motion and communication. University of Pennsylvania Press

Bower G (1981) Mood and memory. Am Psychol 36(2)

Briggs P, Burford B, Dracup C (1998) Modeling self-confidence in users of a computer system showing unrepresentative design. Int J Hum Comput Stud 49:717–742

Casanueva JS, Blake EH (2001) The effects of avatars on co-presence in a collaborative virtual environment. Technical Report CS01–02–00, Department of Computer Science, University of Cape Town, South Africa

Chovil N (1991) Discourse-oriented facial displays in conversation. Res Lang Soc Interact 25:163–194

Collet C, Vernet-Maury E, Delhomme G, Dittmar A (1997) Autonomic nervous system response patterns specificity to basic emotions. J Auton Nerv Syst 62(1–2):45–57

Colquitt JA, LePine JA, Noe RA (2000) Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. J Appl Psychol 85:678–707

Crist TM, Kaufman SB, Crampton KR (1996) Home telemedicine: a home health care agency strategy for maximizing resources. Home Health Care Manage Pract 8:1-9

Damasio A (1994) Descartes' error. Avon, New York

Darkins AW, Carey MA (2000) Telemedicine and telehealth: principles, policies performance and pitfalls. Springer, Berlin Heidelberg New York

Derryberry D, Tucker D (1992) Neural mechanisms of emotion. J Consult Clin Psychol 60(3):329–337

Dillon C, Keogh E, Freeman J, Davidoff J (2000) Aroused and immersed: the psychophysiology of presence. In: Proceedings of 3rd International Workshop on Presence, Delft University of Technology, Delft, The Netherlands, March 2000, pp 27–28

Ekman P (1989) Handbook of social psychophysiology, pp 143–146. Wiley, Chichester

Ekman P, Levenson RW, Friesen WV (1983) Autonomic nervous system activity distinguishes between emotions. Science 221:1208–1210

Ekman P, Friesen WV (1975) Unmasking the face: a guide to recognizing emotions from facial expressions. Prentice Hall, Englewood Cliffs, NJ

Feldman Barrett L, Gross JJ, Conner Christensen T, Benvenuto M (2001) Knowing what you're feeling and knowing what to do

about it: mapping the relation between emotion differentiation and emotion regulation. Cognit Emotion 15:713–724
Frijda N (1986) The emotions. New York: Cambridge University Press. MIT Press
Goleman D (1995) Emotional intelligence. Bantam, New York
Gross JJ, Levenson RW (1997) Hiding feelings: the acute effects of inhibiting negative and positive emotions. J Abnorm Psychol 10(1):95–103
Gross JJ, Levenson RW (1995) Emotion elicitation using films. Cognit Emotion 9:87–108
Guinn C, Hubal H (2003) Extracting emotional information from the text of spoken dialog. In: Proceedings of user modeling (UM) 03 Workshop "assessing and adapting to user attitudes and affect: why, when and how?", Pittsburgh, PA
Hagan MT, Menhaj MB (1994) Training feedforward networks with the marquardt algorithm. IEEE Trans Neural Netw 5(6): 989–993
IJsselsteijn WA (2002) Elements of a multi-level theory of presence: phenomenology, mental processing and neural correlates. In: Proceedings of PRESENCE 2002, pp 245–259 Universidade Fernando Pessoa, Porto, Portugal, 9–11 October 2002
IJsselsteijn WA, de Ridder H, Freeman J, Avons SE (2000) Presence: concept, determinants and measurement. In: Proceedings of the SPIE, Human Vision and Electronic Imaging V, 3959–76
James L (2000) Road rage and aggressive driving. Prometheus, Amherst, NY
Kalawsky RS (2000) The validity of presence as a reliable human performance metric in immersive environments. In: Proceedings of 3rd international workshop on presence, Delft University of Technology, Delft, The Netherlands
Larson J, Rodriguez C (1999) Road rage to road-wise. Tom Doherty Associates, New York
Lewis VE, Williams RN (1989) Mood-congruent vs mood-state-dependent learning: implications for a view of emotion. J Soc Behav Pers 4:157–171
Ledoux J (1992) Brain mechanisms of emotion and emotional learning. Curr Opin Neurobiol 2:191–197
Lisetti CL (1999) A user model of emotion-cognition. In Proceedings of the UM'99 Workshop on Attitude, Personality, and Emotions in User-Adapted Interaction (Banff, Canada, June 1999)
Lisetti CL, Nasoz F (2002) MAUI: a multimodal affective user interface. In: Proceedings of ACM Multimedia International Conference, Juan les Pins, France, December 2002
Lisetti CL, Nasoz F, Lerouge C, Ozyer O, Alvarez K (2003) Developing multimodal intelligent affective interfaces for tele-home health care. Int J Hum Comput Stud 59(1–2):245–255
Lombard M, Ditton T (1997) At the heart of it all: the concept of presence. J Comput Mediated Commun 3(2)
Lorenz R, Gregory RP, Davis DL (2000) Utility of a brief self-efficacy scale in clinical training program evaluation. Eval Health Prof 23:182–193
Martocchio JJ, Dulebohn J (1994) Performance feedback effects in training: the role of perceived controllability. Pers Psychol 47:357–373
Martocchio JJ, Judge TA (1997) Relationship between conscientiousness and learning in employee training: mediating influences of self-deception and self-efficacy. J Appl Psychol 82:764–773
Martocchio JJ (1994) Effects of conceptions of ability on anxiety, self-efficacy, and learning in training. J Appl Psychol 79:819–825
Minsky M (1980) Telepresence. Omni, June 1980:45–51
Mitchell TM (1997) Machine learning. McGraw-Hill
Nicol AA (1999) Presenting your findings: a practical guide for creating tables. American Physiological Association, Washington, DC
Picard RW, Healey J, Vyzas E (2001) Toward machine emotional intelligence analysis of affective physiological state. IEEE Trans Pattern Anal 23(10):1175–1191
Rozell EJ, Gardner WL (2000) Cognitive, motivation, and affective processes associated with computer-related performance: a path analysis. Comput Hum Behav 16:199–222
Takeuchi A, Nagao K (1993) Communicative facial displays as a new conversational modality. In: Proceedings of the INTERCHI'93 conference on human factors in computing systems, Amsterdam pp 187–193
Warner I (1997) Telemedicine applications for home health care. J Telemed Telecare 3:65–66
Warr P, Bunce D (1995) Trainee characteristics and the outcomes of open learning. Pers Psychol 48:347–375
Zajonc R (1984) On the primacy of affect. Am Psychol 39:117–124