



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Int. J. Human-Computer Studies 59 (2003) 245–255

International Journal of
Human-Computer
Studies

www.elsevier.com/locate/ijhcs

Developing multimodal intelligent affective interfaces for tele-home health care

C. Lisetti^{a,*}, F. Nasoz^a, C. LeRouge^b, O. Ozyer^a, K. Alvarez^c

^a*Department of Computer Science, University of Central Florida, Orlando, FL 32816-2362, USA*

^b*Department of Information Systems, University of South Florida, Tampa, FL 33620, USA*

^c*Department of Psychology, University of Central Florida, Orlando, FL 32816, USA*

Received 2 December 2002; received in revised form 16 January 2003; accepted 20 January 2003

Abstract

Accounting for a patient's emotional state is integral in medical care. Tele-health research attests to the challenge clinicians must overcome in assessing patient emotional state when modalities are limited (J. Adv. Nurs. 36(5) 668). The extra effort involved in addressing this challenge requires attention, skill, and time. Large caseloads may not afford tele-home health-care (tele-HHC) clinicians the time and focus necessary to accurately assess emotional states and trends. Unstructured interviews with experienced tele-HHC providers support the introduction of objective indicators of patients' emotional status in a useful form to enhance patient care. We discuss our contribution to addressing this challenge, which involves building user models not only of the physical characteristics of users—in our case patients—but also models of their emotions. We explain our research in progress on Affective Computing for tele-HHC applications, which includes: developing a system architecture for monitoring and responding to human multimodal affect and emotions via multimedia and empathetic avatars; mapping of physiological signals to emotions and synthesizing the patient's affective information for the health-care provider. Our results using a wireless non-invasive wearable computer to collect physiological signals and mapping these to emotional states show the feasibility of our approach, for which we lastly discuss the future research issues that we have identified.

© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: User modeling; Intelligent user interfaces; Emotions; Affective computing; Tele-home health care; Tele-health; Human factors of multimedia systems

*Corresponding author.

1. Introduction

With the mass appeal of Internet-centered applications, it has become obvious that the digital computer is no longer viewed as a machine whose main purpose is to compute, but rather as a machine (with its attendant peripherals and networks) that provides new ways for human–computer interaction (HCI, henceforth) and for computer-mediated communication (CMC, henceforth) among users. Indeed, computers and robots are rapidly entering areas of our lives that typically involve *socio-emotional content*,¹ such as telephone computerized receptionist, service robots in hospitals, homes, and offices, internet-based patient advising (where patients read textual information about their diseases), internet-based health chat lines (most often used for private mental health patient–clinician communications), and computer-mediated patient monitoring and caring. Many similar applications are in the making.

In this article, we focus on tele-home health care (tele-HHC, henceforth) whose viable presence in the United States dates from the 1990s. Tele-HHC provides communication options between the medical professional and patient when hands-on care is not required. For example, tele-HHC interventions are currently used to collect vital sign data remotely (e.g. ECG, blood pressure, oxygen saturation, heart and breath sounds), verify compliance with medicine regimes, improve diet compliance, and assess mental or emotional status (Allen and Roman, 1996; Crist et al., 1996; Warner, 1997; Darkins and Carey, 2000; Johnston et al., 2000). With increasing use of tele-home health care, it is therefore important that the caregiver and care recipient communicate along the affective channel to provide better assessment and responsiveness. However, formulating an assessment may be particularly difficult in tele-HHC settings where patients are treated and monitored remotely by medical professionals using multiple media devices with cues filtered out. Social presence during patient–physician communication is indeed essential, and the rising emergence of tele-home health approaches render research efforts aimed at enhancing such presence urgently necessary. Furthermore, not only may appropriate emotional state assessment be a key indicator of the patient’s mental or physical health status, but the power of emotions themselves over the recovery process has also been documented (Damasio, 1994).

Unstructured interviews with tele-health care providers employed by two tele-home health leaders support the introduction of objective indicators of patients’ emotional status in a usable form to greatly enhance patient care. We discuss our contribution to meeting this challenge, which involves building user models of the physical characteristics of users—in our case patients—and models of their emotions. In this article, we describe our research in progress on Affective Computing for tele-HHC applications which includes: developing a system architecture for monitoring and responding to patients multimodal affect and

¹Socio-emotional content is defined as interactions that show solidarity, compassion, tension relief, agreement, antagonism, tension, and disagreement, in contrast with task-dimensional content which is defined as interactions that ask for or give factual information or opinions.

emotions via multimedia and empathetic avatars (Section 2); developing algorithms to map physiological signals to emotions, and synthesizing the patient’s affective information for use by the health-care provider (Section 3). We conclude by discussing research issues and future research directions (Section 4).

2. MOUE: building models of user’s emotions by sensing, interpreting and adapting to multimodal affective expressions

MOUE (a Model Of User’s Emotions) is a system we developed which builds a model of user’s emotions by first observing the user (e.g. patient) via multi-sensory devices: camera, mouse, keyboard, microphone and wearable computer. The overall paradigm for MOUE (Bianchi and Lisetti, 2002) is shown in Fig. 1, whereas Fig. 2 shows the wearable computer we use in conjunction to existing tele-HHC multimedia technology (shown in Fig. 3) to adapt MOUE to tele-HHC needs. Adaptations of MOUE (discussed later) is a step into the direction of addressing some of the tele-HHC challenges regarding the need for clinician–patient communication along the affective channel. Our MOUE system currently under development for a variety of applications aims at:

- Identifying emotion components from the system sensory observations.

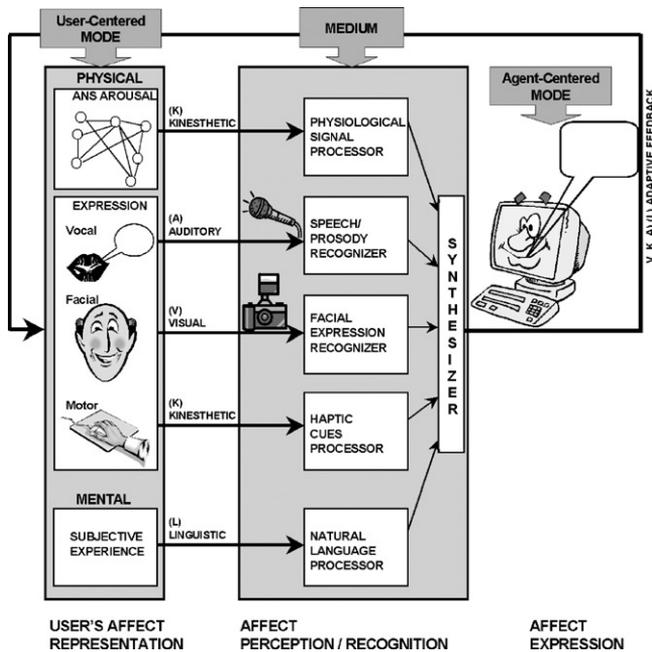


Fig. 1. MOUE paradigm: combining AI for effective affective HCI or CMC.



Fig. 2. Wearable computer bodymedia sensewear.

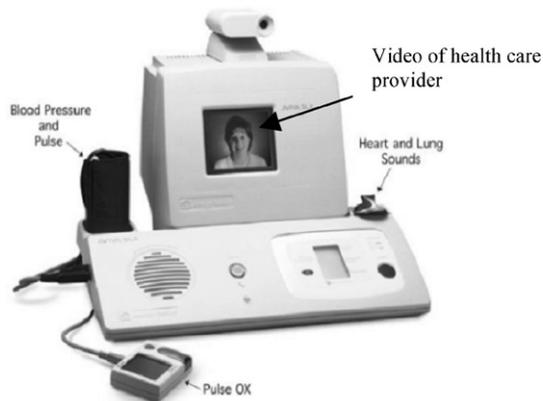


Fig. 3. Video tele-health system.

- Having a database of emotion concepts for each of the possible emotions experienced by a given user.
- Categorizing similar emotions and generating statistical data on user's recurring affective states.
- Inferring emotional trends over time and offering predictions about the user's next affective state.
- Providing feedback to the user about his or her state.
- Adapting to the user's states dynamically via a variety of appropriate means depending upon the given context.

MOUE (Bianchi and Lisetti, 2002) has three main components: (1) a sensory apparatus identifying the most likely emotion(s) experienced by the user during HCI and CMC; (2) an ontology of emotion concepts representing detailed characteristics of the current emotion; and (3) an active interface which externalizes and adapts to the user's perceived emotional state via a variety of modes depending upon the current context.

2.1. MOUE input

Emotions are typically associated with observable physiological changing patterns as well as subjective appraisals and cognitive components representation (Ekman et al., 1983; Zajonc and Markus, 1984; Frijda, 1986; Ortony et al., 1988; Levenson et al., 1990). Therefore, in order to make emotion recognition accurate and reliable, our system takes as input *both* mental subjective components (via text or voice) *and* physiological components (currently facial expressions, body temperature, galvanic skin response and heart rate) associated with emotions experienced by the user. Using the same terminology introduced by Maybury and Wahlster (1998) (see block arrows in Fig. 1), physiological components are identified and collected by associating the different human *modalities* or *modes* employed to express emotion—Visual (facial expression) (Lisetti and Schiano, 2000), Kinesthetic (physiological signals), and Auditory (vocal intonation) (V,K,A)²—with multimodal sensors. The system can also receive input from Linguistic tools (L) in the form of linguistic terms for emotion concepts, which describe the subjective experience associated with a particular emotion, or the emotion label itself. Hence, the system observes the user via specific sensors or *media* associated with specific modalities (see Figs. 1–3): camera, mouse, wearable computer, and microphone. Each medium input is processed using a variety of pattern recognition algorithms (see Section 3) and these various recognized signals are then combined and synthesized into a computational representation that corresponds (with high probabilities) to the emotion experienced.

2.2. MOUE output

The current prototype interprets the user data (facial expression and physiological signals) and identifies the user's emotional state in terms of several available categories (happy, sad, frustrated, angry, afraid) along with the emotion components associated with that specific emotion (see Fig. 4, upper left). The choice of components (described in Lisetti, 2002) was performed by intersecting a combination of current emotion theories in order to obtain emotion concepts that are acceptable across theories and useful for clinicians or users wanting to learn about their own emotions. In short, each emotion is associated with a collection of internal components (both mental and physical), a causal chain of events that describes how it was aroused, and an action tendency that points to the functional aspect of the emotional state. This synthesis constitutes a descriptive feedback to the user about his and her current state, derived from a selected sequence of still images from the user's ongoing video stream (lower left) and physiological signals collected via a wireless wearable computer (Fig. 2). Although different contexts and applications might not require the same type of output, a user wanting to notice or better

² Even though the auditory modality is shown in our MOUE paradigm because it is an important mode through which emotions are expressed, we have not yet implemented speech tone or vocal intonation recognition.



Fig. 4. MOUE's multimodal interface.

understand and recognize her/his own affective experience ongoingly during HCI could use the system “as is”.

2.3. MOUE avatar

Using agent-centered *modalities* or *modes*, multimodal feedback can also be given to the user (see Fig. 1)—the choice of mode depending upon the context and application. For example, whereas an interface agent for a car could automatically adjust the radio station or roll down the windows if the driver is falling asleep, an interface agent for an internet-based psychotherapy session can display empathy via an anthropomorphic avatar who adjusts its facial expressions and vocal intonation to the therapist's textual information (see Fig. 4 upper right). Our avatar's functionality is three-fold: it can be used to assist the user in understanding his/her emotional state by prompting the user with simple questions, comparing the various components of the states he/she believes to be in with the system's output (since self-report is often self-misleading); it can be used to mirror the user's emotions with facial expressions to confirm the user's emotions (Fig. 5); and it can be



Fig. 5. Mirroring user's affective states with empathetic avatars.

used to animate a previously text-only internet-based chat session showing empathic expressions (see also deRosis et al. (this issue) for an elaborate approach to animating affective avatars).

We created our Avatar using Haptik PeoplePutty software (Haptik, 2002). While designing the avatar, various features can be changed: its face, skin type (indicating age), skin color, voice, hair, make-up, accessories, background, and most importantly its facial expression displaying “its emotional state”. We currently can encode expressions for sad, happy, surprised, and angry, each with varying degrees of intensity. Given that people have varying preferences for the “look and feel” of their interlocutor (be it a physician, nurse, tutor, etc.) we chose a flexible technology which has enabled us to design and build a library of different avatars for people of various ethnic background, gender, and age (we currently have created a collection of 15 different avatars of various ages, gender and ethnic backgrounds). Users can therefore select to interact with an anthropomorphic interface agent of their choice by selecting them from a menu (Fig. 4 upper right). Since MOUE also has face recognition abilities, the system remembers the user's selection from one session to the other, and sets its default value to the recognized user's preferred avatar in the next session. The choice of appropriate system feedback (e.g. no avatar, female/male avatar, young/mature avatar, etc.) will be addressed by a cross gender/racial/age forthcoming experiment.

3. MOUE: mapping physiological signals to emotional states

Since automatic assessment of patient's affective state is a critical component of the tele-HHC approach (Pettinari and Jessopp, 2001), we now describe empirical results of our work in measuring autonomic nervous system (ANS) arousal and physiological signals during emotion experience.

3.1. Experiment

In order to measure physiological signals, we used BodyMedia SenseWear (shown in Fig. 2), a wireless wearable computer capable of measuring galvanic skin response, skin temperature, ambient temperature, heat flow, and movement. Since it is

wireless, it can easily and efficiently be used in real-life scenarios to collect sensory data without distracting the user (Body Media Incorporated, 2002).

We have collected 10 records of physiological data from 10 different subjects. Each record consists of 35 min of data recorded with the SenseWear while eliciting five emotions—neutrality, anger, fear, sadness and frustration. We segmented the data, according to the emotions elicited at corresponding time slots, and then we stored the data in a three-dimensional array. The three dimensions are: the subjects who participated in the experiment, the emotion classes, and the data signal types (GSR, temperature, and heart rate). Each slot of the array consists of the normalized average value of one specific data signal belonging to one specific participant while s/he was experiencing one specific emotion. The values of each data signal type are normalized by using the average value of corresponding data type we collected during the relaxation period of the same participant.

We used two different algorithms in order to analyse the data we stored in our three-dimensional array: (a) the k -nearest-neighbor algorithm and (b) the discriminant function analysis, both of which gave similar results while classifying the emotions. Fig. 6 shows the emotional state changes of all the participants: the x -axis represents the emotions elicited during our study, and the y -axis represents the values in percentages we obtained with our classifying algorithm. As can be observed, when we elicited a sadness for example, our system classified the measured data as 90% sadness (10% fear). In short, our preliminary results could classify anger at with 80% success, fear with 80%, sadness with 90%, and frustration with 70%. Fig. 7 shows the data collected for a single-subject and plots the emotion interpreted by our algorithm in percentages (y -axis) over time (x -axis). At time t_3 , for example, the system detects that the subject is 60% fearful, whereas at time t_4 it detects sadness. Although these results are preliminary, they indicate the feasibility of mapping physiological signals to emotions using wearable computers. We are currently running extensive experiments on larger number of subjects to confirm our results and refine our algorithms.

3.2. Presenting emotional readings to the health-provider

Our Tele-HCC MOUE System consists of two sub-systems communicating with each other: one is located in patient's home collecting data as previously described

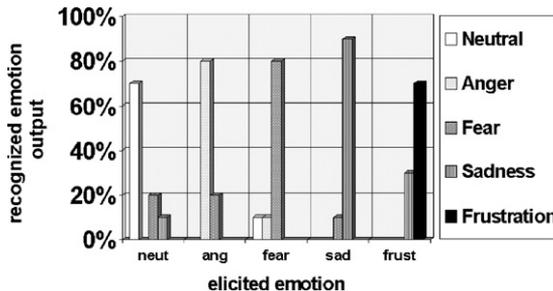


Fig. 6. Physiological signals mapped to emotions.

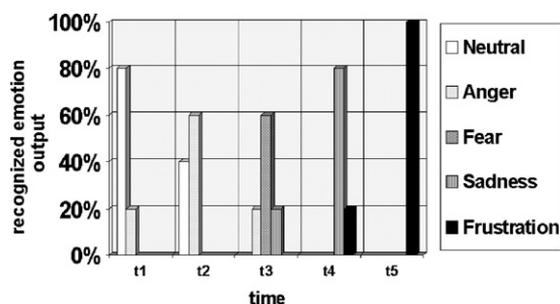


Fig. 7. Emotions experienced over time.

and shown in Fig. 3, and the other one in the health-provider's office, which we now briefly explain. To design an efficient system on the health-provider's side, we conducted semi-structured interviews with tele-home health-providers which revealed that they spend much of their time processing and assimilating data from multiple indicators of patient status in a "manage by exception manner" (Crist et al., 1996). Exceptions are heeded to improve early detection, diagnosis, and treatment. Therefore, information that is displayed in a way that efficiently exploits the strength of the human visual and/or audio system and which is compatible to the health-care provider's mental representation of medical information help the provider's ability to make the correct decisions quickly.

We addressed the providers' concern for interfaces that support process efficiency (i.e. data as necessary), attention to variable changes, and integration of output from various biochemical and affective indicators, by conveying our results in chart form or as alerts (visual or audio) indicating deviations from predetermined standards (Crist et al., 1996). As a result, we proposed to follow the same format as existing reports used by leasing care providers, in which data from an entire patient caseload is integrated on screens that present multiple key indicators for each patient as color-coded key words with drill down options for patient history form, or graphical trend analysis. The chart we proposed (Fig. 7) shows emotional states of a subject at various times. Furthermore, the patient's affective risk level can be computed by summing the probabilities of negative emotions, and highlighting any entry above 80% (not shown). In practice, any of the high-risk flags can indicate the need for further attention, such as a home visit by a clinician.

4. Discussions and conclusions

Currently MOUE is progressing towards detecting and responding to patient's ongoing multimodal emotional states via multimedia. While developing our system, we have become aware of some inherent difficulties in recognizing human emotions

from physiological signals, which we will address in our future research outlined below:

- (1) Design studies that elicit moderate degrees of emotional intensity in single elicitations for short periods of time in order to more confidently attribute increases in arousal to the emotion rather than erroneous artifacts (e.g. compensation mechanisms, movement, habituation, fatigue, irrelevant emotional states; [Levenson, 1988](#)).
- (2) Gather baseline data that involves a moderate degree of physiological arousal in order to (a) capture emotional-physiological events in a more naturalistic way and (b) examine decreases in arousal without a floor limit ([Levenson, 1988](#)).
- (3) Subtract baseline information and/or no emotion variance from emotion data in order to clean up the problems associated with variations in mood and physiology, especially when collecting data from the same subjects over several days ([Picard et al., 2001](#)).
- (4) Because the stability of ANS activity may differ for each physiological response, gather multiple baseline data (e.g. between exercises) and provide ample time for subjects to adjust to the experimental environment and after completing each exercise ([Levenson, 1988](#)).
- (5) Verify that subjects experience the intended emotion(s) via self-report and/or recordings of facial expressions ([Levenson, 1988](#)). The multimedia nature of our system provides the perfect tool for such assessment and verification.
- (6) Because this field is in its infancy, conduct studies and/or keep up to date on new findings regarding the onset and duration of each emotion in order to ensure that (a) physiological data is gathered at the correct time and (b) aggregated data does not include irrelevant time periods ([Levenson, 1988](#)).
- (7) In order to further the field and measure appropriate elements of physiological events, conduct studies and/or keep up to date on (a) the relationship(s) between physiological and psychological events and (b) whether these relationships are limited (e.g. context bound) or can be generalized ([Cacioppo and Tassinari, 1990](#)).
- (8) In order to control for measurement artifacts, standardize the methods used for hand washing, movement, gel application, and placement of electrodes ([Picard et al., 2001](#)).

We also aim at building a tool that can infer patterns, predict, and make suggestions to the patient as to how to move from one emotional state to another. Integrating MOUE with expert systems that could collectively assess health indicators and patterns for diagnostic, predictive and prescriptive purposes is also on our research agenda.

References

- Allen, A., Roman, L., et al., 1996. Home health visits using a cable television network: user satisfaction. *Journal of Telemedicine and Telecare* 2, 92–94.

- Bianchi, N., Lisetti, C.L., 2002. Modeling multimodal expression of user's affective subjective experience. *User Modeling and User Adapted Interaction* 12 (1), 49–84.
- Body Media Incorporated, 2002. www.bodymedia.com
- Cacioppo, J.T., Tassinary, L.G., 1990. Inferring psychological significance from physiological signals. *American Psychologist* 45, 16–28.
- Crist, T.M., Kaufman, S.B., Crampton, K.R., 1996. Home telemedicine: a home health care agency strategy for maximizing resources. *Home Health Care Manage Practice* 8, 1–9.
- Damasio, A., 1994. *Descartes' Error*. Avon Books, New York, NY.
- Darkins, A.W., Carey, M.A., 2000. *Telemedicine and telehealth: principles, policies performance and pitfalls*. Springer Publishing Company, Inc., New York, NY.
- Ekman, P., Levenson, R.W., Friesen, W.V., 1983. Autonomic nervous system activity distinguishes between emotions. *Science* 221 (4616), 1208–1210.
- Fridja, N., 1986. *The Emotions*. Cambridge University Press, New York, NY.
- HapteK, 2002. www.hapteK.com
- Johnston, B., Wheeler, L., Deuser, J., Sousa, K.H., 2000. Outcomes of the Kaiser permanent tele-home health research project. *Archives of Family Medicine* 9, 40–45.
- Levenson, R.W., 1988. Emotion and the autonomic nervous system: a prospectus for research on autonomic specificity. In: Wagner, H.L. (Ed.), *Social Psychophysiology and Emotion: Theory and Clinical Application*. Wiley, Hoboken, NJ, pp. 17–42.
- Levenson, R.W., Ekman, P., Friesen, W.V., 1990. Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology* 27, 363–384.
- Lisetti, C.L., 2002. Personality, affect and emotion taxonomy for socially intelligent agents. In: *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference (FLAIRS'02)*. Pensacola, FL, May 2002, AAAI Press, Menlo Park, CA.
- Lisetti, C.L., Schiano, D., 2000. Automatic facial expression interpretation. *Pragmatics and Cognition* 8 (1), 185–1235.
- Maybury, M., Wahlster, W. (Eds.), 1998. *Readings in Intelligent User Interfaces*. Morgan Kaufman Press, Los Altos, CA.
- Ortony, A., Clore, G., Collins, A., 1988. *The Cognitive Structure of Emotions*. Cambridge University Press, New York, NY.
- Pettinari, C.J., Jessopp, L., 2001. Your ears become your eyes: managing the absence of visibility in NHS Direct. *Journal of Advanced Nursing* 36 (5), 668–675.
- Picard, R.W., Vizas, E., Healey, J., 2001. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 1175–1191.
- Warner, I., 1997. Telemedicine applications for home health care. *Journal of Telemedicine and Telecare* 3, 65–66.
- Zajonc, M., Markus, H., 1984. Affect and cognition: the hard interface. In: Izard, J., et al. (Eds.), *Emotion, Cognition and Behavior*. Cambridge University Press, Cambridge, pp. 73–102.