A Study on Using Biometric Sensors for Monitoring User Emotions in Educational Games.

Cristina Conati, Romain Chabbal, Heather Maclaren
Department of Computer Science
University of British Columbia
2366 Main Mall
Vancouver, BC, V6T 1Z4, Canada
{conati, maclaren}@cs.ubc.ca

Abstract

We present a study designed to assess the performance of biometric signals in detecting user affective states of students interacting with an educational game. We report observations from the study, and preliminary results from the data analysis.

1. Introduction

Assessing the emotions of a user within a real-time environment is an important aspect of affective computing (Picard, 1995) that is currently receiving a substantial level of exploration. The assumption is that, with information on the user's affective state, a system can interact more effectively with the user. How detailed this information should be depends on the task and the application, going from overall valence and level of arousal of the user emotional state (Ball & Breeze, 1999), to a single specific emotion (e.g. emotional stress (Hudlicka & McNeese, 2002; Healey, 2000)), to multiple specific emotions (Picard, Vyzas & Healey, 2001; Conati & Zhou, 2002; Zhou & Conati, 2003). Clearly, the more detailed the emotional information required, the more uncertain and prone to error the modeling task becomes.

The long-term goal of our research is to devise a framework for affective modeling that can detect multiple specific emotions (possibly occurring concurrently), and allow an intelligent interactive agent to use this information to tailor its responses to the user's needs. To handle the high level of uncertainty in this modeling task, the framework is designed to integrate in a Dynamic Decision Network (Dean & Kanazawa, 1989) information on both the *causes* of a user's emotional reactions and their *effects* on the user's bodily expressions (Conati, 2002). In this paper we present an empirical study designed to understand how these effects can be monitored in real-time by using affective biometric signals (Picard, 1995).

To date, the assessment of emotions from biometric signals within a real-time environment has been attempted only in quite constrained contexts such as the detection of a specific emotion (level of anxiety), in a driver (Healey, 2000) and in combat pilots (Hudlicka & McNeese, 2002). Similarly, Scheirer et al. have used biometric sensors to monitor the induced frustration of users (Scheirer, Fernandex, Klein, & Picard, 2002). Biometric sensors have been used to detect several different emotions only when the emotions were deliberately expressed by a professional actress (Picard, Vyzas, & Healey, 2001).

Using biometric sensors for real-time recognition of multiple emotions, expressed spontaneously in an uncontrolled environment, has not been previously attempted. Thus, it is not clear how effectively the sensors can detect emotions that may be expressed more subtly, that could interfere with each other and whose duration is unknown, in an environment with a possibly higher level of noise due to motion artefacts.

To shed light on this issue, we have conducted a user study in which four biometric sensors were used to record the emotions expressed by students that were playing Prime Climb, an electronic educational game.

In this paper, we give an overview of the study performed and illustrate the difficulties encountered due to operation in an uncontrolled environment. We discuss the observations made of the monitoring process itself, and present initial findings whilst attempting to process the collected data in a real-time manner.

2. Monitoring Emotional Response in Educational Games

As mentioned earlier, how much information on the user affect a system needs depends on how useful this information can be when tailoring the interaction. Educational games are examples of applications in which the more the system knows about the user's detailed emotions, the better it can react by tailoring the game so that it promotes learning while maintaining a high level of engagement. The range of emotions that educational games can induce is very wide, and depends on each player's personality and learning style, especially in multi-player games (Conati & Klawe, 2002).

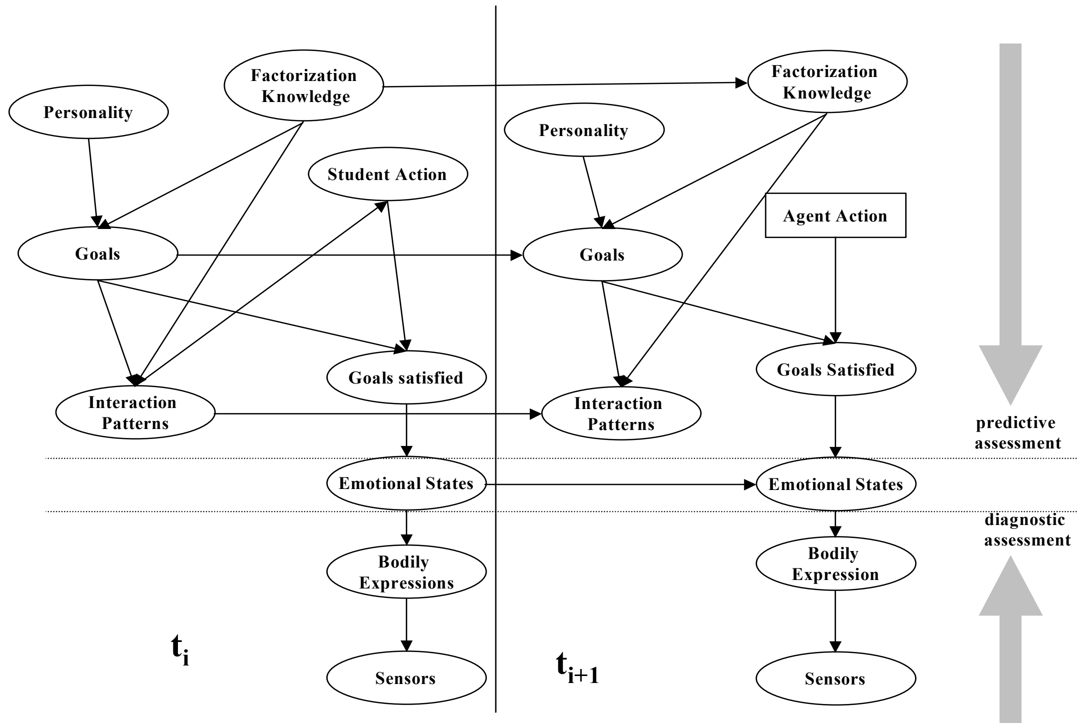

Figure 1. Two time slices of the DDN model of user affect

Thus, we are using the educational game, Prime Climb, developed by the EGEMS group at the University of British Columbia (UBC) to help 6th and 7th grade students learn number factorization, as the test bed for our research on modeling detailed user emotions.

In the game, two players must cooperate to climb a series of mountains that are divided into numbered sectors. Each player can only move to a numbered sector that does not share any factors with the sector occupied by his partner. Figure 2 shows a screen shot of the Prime Climb interface. When a player makes a wrong move, she falls and starts swinging from the climbing rope. Some tools are available to help the players decide which sector to move to, including a pedagogical agent that will act as an advisor. One phase of Prime Climb, called "Practice Climb", involves a single player controlling one climber and the agent controlling the other climber. The goal of the pedagogical agent is to provide tailored support, to help the player to learn number factorization while maintaining a high level of engagement. Thus, enabling the agent to identify specific emotions experienced by a player will allow it to identify the form of feedback that will produce the most beneficial response at that point in the game.

As mentioned earlier, our approach to enable the agent to use this form of reasoning is to combine in a Dynamic Decision Network (Dean & Kanazawa, 1989) information from possible causes of emotional arousal (*predictive assessment* in Figure 1) with information on relevant user bodily expressions detected through appropriate sensors (*diagnostic assessment* in Figure 1). Following the OCC cognitive theory of emotions (Ortony, Clore, & Collins, 1988), the causal part of the network encodes how the interaction of user's goals with the current state of the game (nodes *student action outcome* and *agent action outcome* in Figure 1) influences the user's emotional state, (Conati & Zhou, 2002; Zhou & Conati, 2003), represented in terms of 6

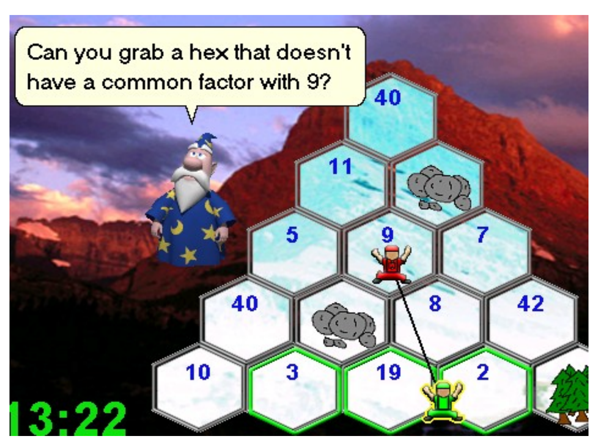

Figure 2. The Prime Climb educational game

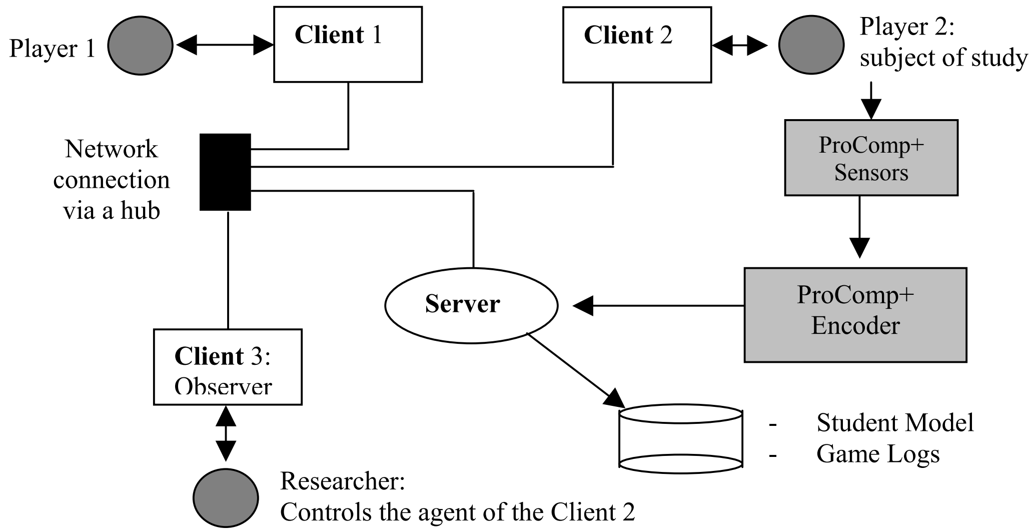

Figure 3. Setting of a Prime Climb session

of the 22 emotions defined by the OCC theory: *joy, distress, admiration, reproach, pride* and *shame*.

The study we next describe was conducted to collect data to define the diagnostic part of the network, i.e. the structure between the nodes representing user emotions and sensorial evidence. We focused on four of the biometric signals commonly used for emotion detection (Picard, 1995): skin-conductivity (SC), electromyography (EMG) of the muscles involved in frowning and corrugating eyebrows, blood-volume pressure (BVP), and respiration (RESP).

3. GEERing up! Prime Climb

The user study was carried out during the UBC GEERing up! summer camp, which involves children from grade 2 to 10 in various scientific and engineering activities.

The study objective was to ascertain how much information on user affect the four sensors could provide in an uncontrolled setting in which GEERing up! students interacted freely with Prime Climb. This included whether the four sensors could identify the six specific emotions currently modelled in our dynamic decision network, or whether only levels of valence and arousal could be reliably retrieved.

The hardware for the experiment was set up as shown in Figure 3. During each game session one player was monitored using sensors for the four physiological signals described earlier. In some of the sessions, the player climbed with the agent only (controlled by an experimenter through a Wizard of Oz interface), to provide data for the interaction currently modelled by our dynamic decision network. In the rest of the sessions, the player climbed with another student, while the agent still provided aid to the player being monitored. Although we still do not have an affective model for this more complicated interaction, in which a players emotional reactions toward both the climbing peer and the agent need to be assessed, we wanted to collect data for both situations whilst the opportunity was available.

One of the main challenges of creating a mapping between sensor data and user affect in an uncontrolled environment is that the user's true emotional states are not known. They can only be approximately identified through subjective observations of the interaction. In our study, these observations were derived post-study from a time-stamped video recording of each interaction session. Each recording was synchronized with logs of the game events that could stimulate an emotional reaction (e.g., a move, a fall, an agent intervention). Each video data segment was then annotated by a human observer with information on the user's emotional state, including valence (with ratings from "- -", very negative, to "+ +", very positive), arousal (rated from 1 to 5), and a specific emotion when possible. These annotations were then processed by software to extract all the segments corresponding to a given criterion (e.g. arousal > 1), retrieve the sensor data over those segments and plot the results for each segment as shown in Figures 4 and 5. The arousal and valence ratings given at the top of these figures correspond to the ratings given within the observer's annotations for that segment. We currently have complete sensor and labelled video data on 8 subjects.

The first step of data analysis was to inspect the collected data “by hand” to get a general sense of the sensors performance. In the next session we report preliminary results of this phase, after discussing a few lessons we learned from the study.

4. Lessons learned and Preliminary Findings

The first lesson we learned from the study relates to the concern often raised on the intrusiveness of biometric sensors. In our study, we found that not only did the children who wore the sensors not find them intrusive, but those not selected to wear the sensors were often disappointed because they regarded this activity as part of the game. This would suggest that, in a game-like environment aimed at very young people, the concern over intrusiveness could be downgraded.

The second lesson is that reliably assessing the user’s true emotional state in an uncontrolled environment is very difficult. Our video recordings showed that users’ visible bodily expressions often did not give enough indication of what they were experiencing, and therefore video annotations required a high level of subjective judgement. Originally, only one set of annotations was made, which the initial results presented here are based on. An additional reviewer produced a second set of annotations to increase reliability, and we are working on correlating both sets and the sensor data for the next stages of analysis.

The third lesson is that the level of noise in all the signals increases in uncontrolled environments in which subjects have high mobility (as ours did), because all four sensors are sensitive to motion artefacts. RESP and BPV were the most affected, resulting in signals so noisy that no meaningful patterns could be discerned during preliminary analysis. Thus, for the time being we decided to concentrate our efforts on the seemingly less noisy SC and EMG signals, producing the following findings.

1. The data collected from the SC signal seems to confirm previous findings that sudden arousal can cause a startle response in the signal and that the relative amplitude of this pattern is in direct relation to the user’s arousal level. Figure 4a shows an example SC pattern detected for an emotional response with arousal level of 2 and negative valence. However, the overall amplitudes of the patterns recorded were dependent upon the user, suggesting that personalization is required to ensure consistent recognition of significant patterns for this signal.
2. Potential SC patterns (such as the one shown in Figure 4a) were detected by using the gradient of the signal. When a point is detected at which the gradient begins to increase, the successive values are searched for a point at which the gradient of indicates that a peak has been reached. If the amplitude of the signal, defined as the difference in value between these two data points, is above a given threshold, then the pattern is deemed to be significant. Unfortunately, this method is susceptible to noise introduced by motion artefacts (as shown in Figure 4b). Methods to detect and remove these effects are currently under investigation.
3. Communication of emotion to another human player seems to reduce internal arousal response (i.e. the amplitude of any SC pattern recorded at the same time). Two of our subjects were observed to have an unusually low percentage of segments with emotional arousal that corresponded to SC startle responses (13% and 7% respectively). The only detectable difference between these two subjects and the others is that they both talked to their peer player during the game session. This suggests that explicitly expressing emotions to another person could be regarded as ‘releasing the internal emotional pressure’ and should be therefore taken into account when analyzing SC data. However, this statement calls for further investigation.
4. Previous studies (e.g. the work by Ekman, Levenson, & Friesen (1983)) indicated that the amplitude of the SC signal should be higher for emotions with negative valence than for emotions with positive valence. However, this phenomenon was not observed in our data. Given the difference in amplitude between SC recordings for each individual and that differences in amplitude also correspond to different levels of arousal, a more robust method to determine emotional valence is required.
5. The EMG signal indicates muscle contraction through an increase of voltage from a tonic baseline, and it has been used in previous studies to distinguish between a frown (often an indicator of emotions such as puzzlement, anger and frustration) and an eyebrow lift (a possible indicator of surprise). Although the preliminary analysis of this signal in our study showed that both frowns and eyebrow lifts could generate peaks as shown in Figure 5, these peaks did not allow discrimination between these expressions.

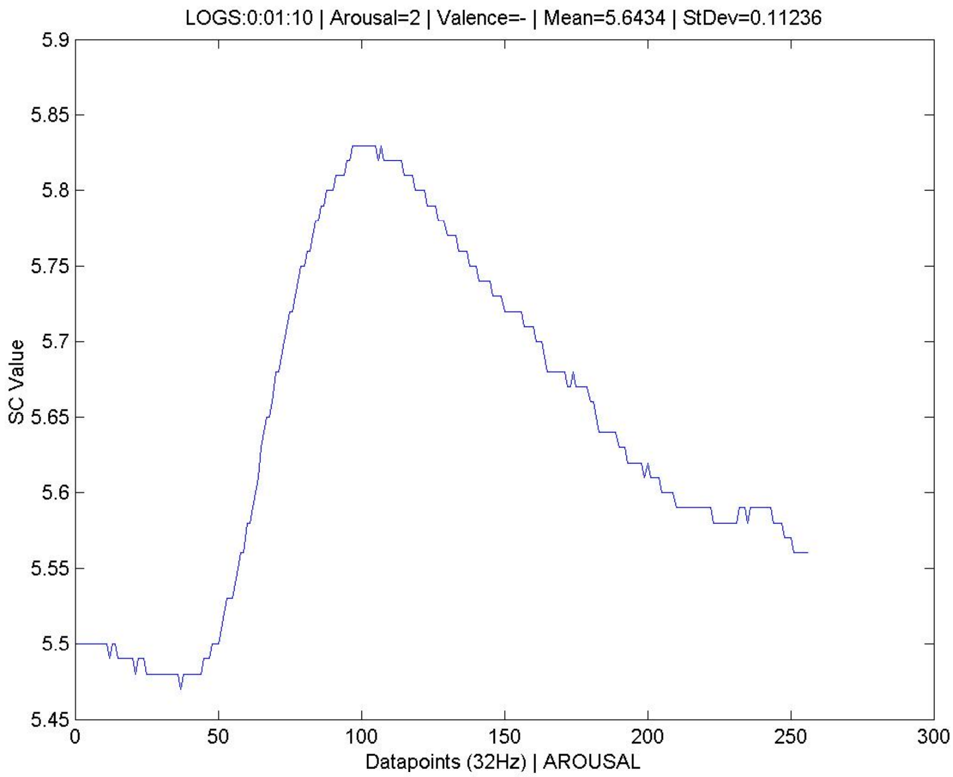

Figure 4a. Plot of a Skin Conductance startle response over one segment of negative emotional arousal

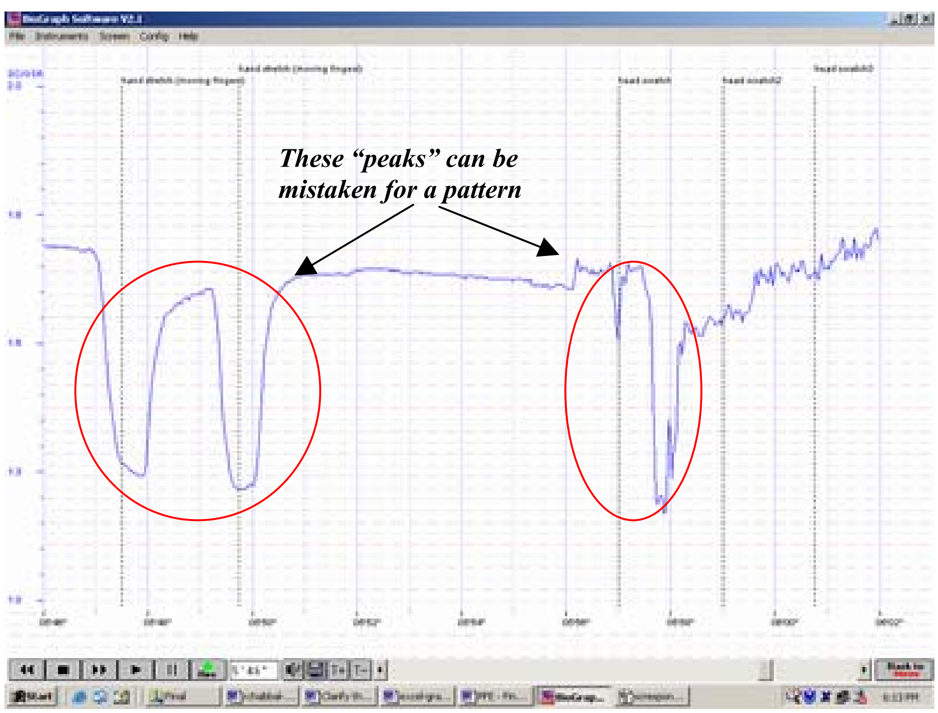

Figure 4b. Alteration of the SC signal upon motion of the left-hand

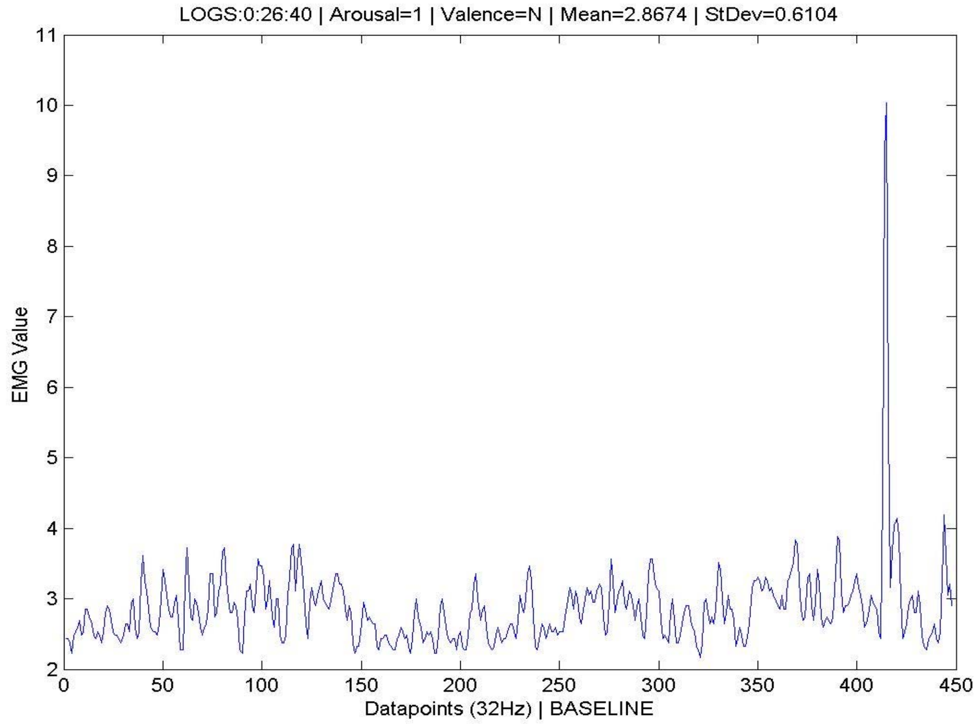

Figure 5. Plot of an EMG peak at the end of a period of rest.

5. Current and Future Work

There are three issues on which our attention is currently focusing. The first relates to the fact that signal patterns must exceed a given threshold in order to be considered significant. However, this threshold is usually user-dependent. Thus, we are investigating a method for threshold selection that starts from a value based on data from a population of previous users and dynamically changes this value, on the basis of time and data collected, to incrementally give more importance to a threshold based on the incoming signal for the current user.

The second issue regards the removal of the SC noise generated by commonly recurring external factors, such as the player moving the appendage to which the sensor is attached.

The final, much larger, issue is the attempted correlation of the annotations from the video data with the patterns detected within the biometric signals. If links can be found between these two sets of data, then this will be a significant step towards detailed emotional analysis in real-time.

6. References

Ball, G. & Breese, J. (1999). Modeling the emotional state of computer users. In *Workshop on 'Attitude, Personality and Emotions in User-Adapted Interaction', UM '99*, Banff, Canada.

Conati, C. (2002). Probabilistic assessment of user's emotions in education games. *Journal of Applied Artificial Intelligence, special issue on " Merging Cognition and Affect in HCI"*, 16 (7-8).

Conati, C. & Zhou, X. (2002). Modeling students' emotions from cognitive appraisal in education games. In *Proceedings of ITS 2002, 6th International Conference on Intelligent Tutoring Systems*, Biarritz, France.

Conati, C. & Klawe, M. (2002). Socially Intelligent Agents in Educational Games. In *Socially Intelligent Agents - Creating Relationships with Computers and Robots.* Dautenhahn, K., Bond A., Canamero, D., & Edmonds B., (Eds). Kluwer Academic Publishers.

Dean, T. & Kanazawa, K. (1989). A model for reasoning about persistence and causation. *Computational Intelligence*, 5(3), 142-150.

Ekman, P., Levenson, W., & Friesen, W.V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221, 1208-1210.

Hudlicka, E. & McNeese, M. (2002). Assessment of User Affective and Belief States for Interface Adaptation: Application to an Air Force Pilot Task. *User Modeling and User Adapted Interaction*, 12(1), 1-47.

Healey, J. (2000). *Wearable and automotive systems for affect recognition from physiology.* PhD thesis, MIT, Cambridge, MA, 2000.

Ortony, A., Clore, G.L., & Collins, A. (1988). *The Cognitive Structure of Emotions*. NY: Cambridge University Press.

Picard, R. (1995). *Affective computing.* Cambridge: MIT Press.

Picard R.W., Vyzas E., & Healey J. (2001). Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10), 1175-1191.

Scheirer, J., Fernandez, R., Klein, J., & Picard R.W. (2002). Frustrating the User on Purpose: A Step Toward Building an Affective Computer. *Interacting with Computers*, 14(2), 93-118.

Zhou, X., & Conati C. (2003). Inferring user goals from personality and behaviour in a causal model of user affect. In *Proceedings of IUI 2003*, Miami, Florida, USA, 211-218.