International Journal of Human-Computer Interaction
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hihc20

Dynamic Difficulty Adjustment in Computer Games Through Real-Time Anxiety-Based Affective Feedback

Changchun Liu [a], Pramila Agrawal [b], Nilanjan Sarkar [a] & Shuo Chen [c]

[a] Vanderbilt University, Nashville, TN, USA
[b] Metis Design Corporation, Cambridge, MA, USA
[c] Emory University, Atlanta, GA, USA
Version of record first published: 04 Aug 2009.

To cite this article: Changchun Liu , Pramila Agrawal , Nilanjan Sarkar & Shuo Chen (2009): Dynamic Difficulty Adjustment in Computer Games Through Real-Time Anxiety-Based Affective Feedback, International Journal of Human-Computer Interaction, 25:6, 506-529

To link to this article: http://dx.doi.org/10.1080/10447310902963944

INTL. JOURNAL OF HUMAN–COMPUTER INTERACTION, *25*(6), 506–529, 2009

ISSN: 1044-7318 print / 1532-7590 online
DOI: 10.1080/10447310902963944

Dynamic Difficulty Adjustment in Computer Games Through Real-Time Anxiety-Based Affective Feedback

Changchun Liu[1], Pramila Agrawal[2], Nilanjan Sarkar[1], and Shuo Chen[3]
[1]Vanderbilt University, Nashville, TN
[2]Metis Design Corporation, Cambridge, MA
[3]Emory University, Atlanta, GA

A number of studies in recent years have investigated the dynamic difficulty adjustment (DDA) mechanism in computer games to automatically tailor gaming experience to individual player's characteristics. Although most of these existing works focus on game adaptation based on player's performance, affective state experienced by the players could play a key role in gaming experience and may provide a useful indicator for a DDA mechanism. In this article, an affect-based DDA was designed and implemented for computer games. In this DDA mechanism, a player's physiological signals were analyzed to infer his or her probable anxiety level, which was chosen as the target affective state, and the game difficulty level was automatically adjusted in real time as a function of the player's affective state. Peripheral physiological signals were measured through wearable biofeedback sensors and several physiological indices were explored to determine their correlations with anxiety. An experimental study was conducted to evaluate the effects of the affect-based DDA on game play by comparing it with a performance-based DDA. This is the first time, that is known, that the impact of a real-time affect-based DDA has been demonstrated experimentally.

1. INTRODUCTION AND MOTIVATION

There has been a steady progress in the field of computer games in recent years that has become one of the most popular and economically successful forms of human–computer interaction (HCI) systems (Zaphiris, 2007). The worldwide market for computer game hardware, software, and accessories is expected to grow from £11.7 billion in 2002 to £17 billion in 2007 (RocResearch, 2004) as more

We gratefully acknowledge the help of Professor Eric Vanman of the Psychology department of Georgia State University during this work.

Correspondence should be addressed to Changchun Liu, Department of Electrical Engineering and Computer Science, Vanderbilt University, VU Station B 351679, 2301 Vanderbilt Place, Nashville, TN 37235-1679. E-mail: changchun.liu@vanderbilt.edu

novel play environments are developed for entertainment and education (Stokes, 2005). Although gaming technology has continued to evolve, there has been general dissatisfaction of players with the current computer games because of their inadequacy of providing optimal challenge levels to accommodate individual player's characteristics such as skills, capacities to learn and adapt, and emotional traits (Gilleade, Dix, & Allanson, 2005; Sweetser & Wyeth, 2005). Static difficulty levels that are manually selected by the players are not sufficient to avoid getting the player overwhelmed or bored because players are likely to be unable to assess which challenge level matches their skills (Koster, 2004). In addition, asking the players to frequently choose the difficulty levels could be annoying as well as cause interruption of the game play (Chen, 2007).

To address this issue, a growing number of studies have been investigating the dynamic difficulty adjustment (DDA) mechanisms to enable the game-playing experiences automatically tailored to the individual characteristics. Demasi and Cruz (2002) developed a "challenge function" by using heuristics (e.g., time to complete a task and rate of successful shots, etc.) to map a given game state to a value that specifies the difficulty felt by a user. Reinforcement learning has been employed to allow computer-controlled agent to learn optimal strategies in a fighting game while choosing suboptimal actions to fit the players' performance when necessary (Andrade, Ramalho, Santana, & Corruble, 2005). Spronck, Ponsen, Sprinkhuizen-Kuyper, and Postma (2006) proposed a rule-based approach, called dynamic scripting, that includes the model of the opponent player. It assigns each behavior rule a probability of being picked and then modifies the probability dynamically based on the success or failure rate of each rule (Spronck et al., 2006). The DDA has been increasingly recognized by the game development community as a key characteristic for a successful game. For instance, in Resident Evil 4 (http://www3.capcom.co.jp/bio4/english.html), a third-person shooter game with five levels of difficulty, the difficulty adjustment can be automatically accomplished based on the player's performance.

In most current DDA research works, the performance of the player has been used as a main measure of the characteristics of the players. However, as noted by Pagulayan, Keeker, Wixon, Romero, and Fuller (2002), unlike productivity software, computer game's paramount evaluation factor should be the affective experience provided by the play environment instead of the user's performance. A case study on several popular computer games (e.g., Combat Flight Simulator (PC), Combat Evolved (Xbox), etc.) suggested that standard performance-based usability methods may not be sufficient to evaluate gaming experience and issues related to affective aspects of the game (e.g., fun) should be considered (Pagulayan, Steury, Fulton, & Romero, 2005). Mandryk and Atkins (2007) also regarded the emotional experience is the key measurement of a game playing and used fuzzy physiological approach to determine the underlying affective states related to game play in an off-line manner. Echoing similar opinion, the concept of "Affective Gaming" has been proposed in recent years (Gilleade & Allanson, 2003; Gilleade et al., 2005; Magerkurth, Cheok, Mandryk, & Nilsen, 2005; Sykes & Brown, 2003), which focuses on exploring the impacts of affective factors in computer game design and adaptation. Furthermore, players can have different motivations to play a game (Koster, 2004). For a player who derives satisfaction

from completing difficult tasks, one has to be cautious in decreasing the difficulty level even when he or she has been defeated for several times, whereas for another player, it may not be appropriate to increase the difficulty level even when his/her performance is excellent. We believe that the affective state of a player is likely to be a critical factor in many gaming experience, and that the next generation of DDA mechanism should consider both player's performance and affective state information.

The primary objective of this research is to explore the feasibility of recognizing a player's affective states via a physiology-based affect recognition technique during gaming and investigate how the gaming experience can be augmented by using the recognized affective state to automatically adjust game difficulty level in real time. Note that we recognize the fact that a DDA mechanism that considers only affective state information may not be optimal. A versatile DDA mechanism should also consider player's performance, his or her personality, and the context and complexity of the game among other issues to generate a rewarding gaming experience. However, we first want to establish that real-time affective feedback is possible during the gaming process and that such a feedback can impact the experience of game play. The goal here is to advance the state of the art in affective gaming, which has gained significant importance in the HCI community in recent years (Gilleade & Allanson, 2003; Gilleade et al., 2005; Magerkurth et al., 2005; Sykes & Brown, 2003). To achieve this objective, we divide our research into two major phases: (a) to obtain the affective model in Phase I, and (b) to investigate the impact of affect-sensitiveness on the gaming experience in Phase II. The primary contribution of this article lies in Phase II work. However, because the Phase II work is dependent on affective models developed in Phase I, we believe it is necessary to briefly discuss Phase I work. The detailed results of the Phase I work were published in (Rani, Liu, Sarkar, & Vanman, 2006) and are omitted here. This is the first time, to our knowledge, that the impact of an affect-based DDA on player's interaction with a computer game that is capable of physiology-based affect recognition and real-time difficulty adjustment in a closed-loop manner has been investigated experimentally.

There are several modalities such as facial expression (Bartlett, Littlewort, Fasel, & Movellan, 2003), vocal intonation (Lee & Narayanan, 2005), gestures (Asha, Ajay, Naznin, George, & Peter, 2005), and physiology (Kulic & Croft, 2007; Leon, Clarke, Callaghan, & Sepulveda, 2004; Mandryk & Atkins, 2007; Rani, Sarkar, Smith, & Kirby, 2004) that can be utilized to recognize the affective states of individuals interacting with a computer. In this work we choose to create affective model based on physiological data for several reasons. One of the chief advantages of using physiology is that physiological signals are continuously available and are not dependent on overt emotional expression. Our aim is to recognize affective states of people engaged in real-life activities, such as playing computer games. Even if a person does not overtly express his or her emotion through speech, gestures, or facial expression, a change in the physiological signal pattern associated with the changes of underlying affective states is likely to occur, which could be detectable. Furthermore, physiology is usually not under one's voluntary control and hence may provide an undiluted assessment of the underlying affective state. It is also reasonably independent of cultural, gender,

and age related biases (Brown, Hall, Holtzer, Brown, & Brown, 1997). Besides, there is evidence that the transition from one affective state to another is accompanied by dynamic shifts in indicators of autonomic nervous system (ANS) activity (Bradley, 2000; Picard, 1997). The physiological signals that have been used in this research consist of various cardiovascular, electrodermal, electromyographic, and body temperature signals, all of which have been extensively investigated in psychophysiology literature (Bradley, 2000).

An important question when estimating human affective state is how to represent the affective state. Although much existing research categorizes human affective states into what is called a set of "basic emotions," there is no consensus on a set of basic emotions among the researchers (Cowie et al., 2001). This fact implies that it requires pragmatic choices to select a target affective state for a given application (Cowie et al., 2001). In this article, we chose anxiety to be the target affective state for the affect-based DDA design. The DDA mechanism will allow the computer game to recognize anxiety and respond to it in an appropriate manner. Anxiety was chosen for two primary reasons. First, anxiety plays an important role in various human-computer interaction tasks that can be related to performance, challenge, and ability (Brown et al., 1997; Chen, 2007). Second, the correlation of anxiety with physiology is well established in the psychophysiology literature (Rohrmann, Hennig, & Netter, 1999) and thus provides us with a scientific basis to infer it. In this study, we develop an affective model of a player that is capable of determining the intensity of anxiety (i.e., low/medium/high) instead of discrete emotions. Another important fact that should be noted for affective modeling is the phenomenon of person stereotypy. There is evidence that within a given context, different individuals express the same emotion with different physiological response patterns (Lacey & Lacey, 1958). The novelty of the presented affective modeling is that it is individual specific to accommodate the differences encountered in emotion expression.

Note that a player's performance and affective state could be fused together in a DDA mechanism. However, in this article we focus on the impact of an affect-based DDA on the gaming experience. Hence, we separated a performance-based DDA from an affect-based DDA and compared their effects on a computer game. In addition, we implemented the DDA mechanism using state-flow diagrams, where the states were represented by a set of predefined difficulty levels. Although it is possible to use a player's affective state information to manipulate game environment settings and agents' behaviors in a moment-by-moment manner, such control often depends on heuristic knowledge and specific genre of a game (Hunicke & Chapman, 2004) and is beyond the scope of this article. However, because most existing computer games have embedded predefined difficulty levels, the presented approach could be integrated with a large class of games.

The rest of the article is organized as follows: The next section reports on related works in physiology-based affect recognition, intelligent tutoring system that used affective cues, and affective gaming. A description of the physiological signals and the features that were derived from these signals for affective modeling are presented in section 3. In section 4, we describe the machine learning algorithm used for detecting affective cues. Section 5 presents experimental designs

for affective model building (Phase I) and evaluation of the effects of the affect-based DDA (Phase II). This is followed by a detailed results and discussion section (section 6). Finally, Section 7 summarizes the contributions of the article and provides future directions of this research.

2. RELATED WORK

The use of physiology as a method to evaluate the affective state has attracted increasing attention in recent years. Multiple physiological measures such as electromyography (EMG), electroencephalography (EEG), and heart rate variability (HRV), have been used jointly to assess stress (Rani, Sims, Brackin, & Sarkar, 2002), workload (Kramer, Sirevaag, & Braune, 1987), and mental effort (Vicente, Thornton, & Moray, 1987). Galvanic Skin Response (GSR), EMG and Electrocardiogram (ECG) have been examined in (Mandryk & Atkins, 2007) to determine the underlying affective states related to game play. Various machine learning techniques including fuzzy logic (Mandryk & Atkins, 2007; Rani et al., 2002), discriminant function analysis (Nasoz, Alvarez, Lisetti, & Finkelstein, 2003), auto-associative neural networks (Leon, Clarke, Callaghan, & Sepulveda, 2007), and support vector machines (Kim, Bang, & Kim, 2004) have been applied to differentiate discrete emotions (e.g., anger, joy, sadness, etc.). In our previous work (Rani et al., 2004), we have shown the relationship between anxiety and several physiological parameters like HRV, facial EMG, GSR, blood pulse volume, and peripheral temperature. Although the existing studies provide valuable supports for the validity of physiology-based affect recognition, the impact on human users when computers respond to recognized affective states (i.e., interact in a closed-loop manner) is still largely unexplored.

In the context of intelligent tutoring system, there have been research efforts that aim at endowing a computerized tutor with the ability to adapt affectively in the teaching-learning process, which would permit a more natural, enjoyable and productive discourse. Conati (2002) proposed a probabilistic model to monitor a user's emotion and engagement during automated tutoring. The affective states of students (i.e., reproach, shame, and joy) were detected by the use of eye brow EMG, GSR, and ECG through a dynamic decision network. The trade-off between engagement and learning was achieved by a utility function that assigned appropriate weights to students' performance and engagement. Prendinger, Mori, and Ishizuka (2005) conducted an experimental study that examined GSR and EMG to investigate the effect of a life-like virtual teacher on the affective state of users under "affective persona" and "nonaffective persona" conditions. Our work differs from those studies in several aspects. First, our work focuses on investigating a DDA mechanism in the context of computer games. Specifically we are interested in evaluating the effects of an affect-based DDA on gaming experience by comparing it with a performance-based DDA though a systematic user study. Second, we identify the varying levels of anxiety instead of determining the occurrence of specific discrete emotions. Determining the intensity of an affective state could be a more challenging problem than differentiating discrete emotional states (Rani et al., 2006). Third, we adopt an individual-specific approach to overcome

person-stereotypy and explore a more comprehensive set of physiological indices. We develop affective model for each individual player with reliable real-time predictions (as described in section 6), whereas works in (Conati, 2002; Prendinger et al., 2005) presented across-individuals approach that did not consider person stereotypy.

Finally, our work falls into a nascent research field of HCI, called Affective Gaming (Gilleade & Allanson, 2003; Gilleade et al., 2005; Magerkurth et al., 2005; Sykes & Brown, 2003), that aims at enhancing gaming experience by adapting the game course to the player's affective state. Although concepts of affective gaming have been discussed for game design in (Gilleade & Allanson, 2003; Gilleade & Dix, 2004), the existing studies have the limitation of lacking systematic experimental investigation, which has been addressed in our work.

3. PHYSIOLOGICAL INDICES FOR RECOGNIZING ANXIETY

There is good evidence that the physiological activity associated with the affective state can be differentiated and systematically organized (Bradley, 2000). The relationships between both electrodermal and cardiovascular activities with anxiety were investigated in (Dawson, Schell, & Filion, 1990; Pagani, Lombardi, & Guzzetti, 1986; Rohrmann et al., 1999; Watts, 1975). When a human being is anxious, it is commonly observed that the parasympathetic activity of his/her heart decreases and the sympathetic activity increases (Pagani et al., 1986). It was also reported that anxiety may cause an increase in skin conductance level (Watts, 1975). Previous research has validated blood pulse volume measured at fingers is sensitive to the stress manipulation and is correlated with self-reported anxiety during the anticipation period (Bloom & Trautt, 1978). Measures of EMG activity of the chosen muscles (e.g., Corrugator Supercilii muscles) were also shown to be strong indicators of anxiety (Ekman & Friesen, 1986). In our work, we used this relationship between physiological response and the underlying affective states to develop an affect-recognition system.

The physiological signals we examined were: features of cardiovascular activity, including interbeat interval, relative pulse volume, pulse transit time, heart sound, and preejection period; electrodermal activity (tonic and phasic response from skin conductance) and EMG activity (from Corrugator Supercilii, Zygomaticus, and upper Trapezius muscles). These signals were selected because they are likely to demonstrate variability as a function of the targeted affective states, as well as they can be measured noninvasively, and are relatively resistant to movement artefacts (Lacey & Lacey, 1958).

Multiple features (as shown in Table 1) were derived for each physiological measure by using various signal processing techniques such as Fourier transform, wavelet transform, adaptive thresholding, and peak detection. Some of these features were described in our previous work (Rani et al., 2004). "Sym" is the power associated with the sympathetic nervous system activity of the heart (in the frequency band 0.04–0.15 Hz). "Para" is the power associated with the parasympathetic nervous system activity of the heart (in the frequency band 0.15–0.4 Hz). "VLF" is the power associated with the Very Low Frequency band

Table 1: Physiological Indices

Physiological Response	*Features Derived*	*Label Used*	*Unit of Measurement*
Cardiac activity	Sympathetic power (from electrocardiogram [ECG])	Sym	Unit/Square second
	Parasympathetic power (from ECG)	Para	Unit/Square second
	Very low frequency power (from ECG)	VLF	Unit/Square second
	Ratio of powers	Sym Para Para VLF Sym VLF	No unit
	Mean Interbeat Interval (IBI)	IBI ECG_{mean}	Milliseconds
	Standard deviation of IBI	IBI ECG_{std}	Standard deviation (no unit)
	Mean amplitude of the peak values of the PPG signal (Photoplethysmogram)	PPG $Peak_{mean}$	Micro volts
	Standard deviation (Std.) of the peak values of the PPG signal	PPG $Peak_{std}$	Standard deviation (no unit)
	Mean pulse transit time	PTT_{mean}	Milliseconds
Heart Sound	Mean of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Mean d3 Mean d4 Mean d5	No unit
	Standard deviation of the 3rd, 4th, and 5th level coefficients of the Daubechies wavelet transform of heart sound signal	Std d3 Std d4 Std d5	No unit
Bioimpedance	Mean preejection period	PEP_{mean}	Milliseconds
	Mean IBI	IBI ICG_{mean}	Milliseconds
Electrodermal activity	Mean tonic activity level	$Tonic_{mean}$	Micro-Siemens
	Slope of tonic activity	$Tonic_{slope}$	Micro-Siemens/ second
	Mean amplitude of skin conductance response (phasic activity)	$Phasic_{mean}$	Micro-Siemens
	Maximum amplitude of skin conductance response (phasic activity)	$Phasic_{max}$	Micro-Siemens
	Rate of phasic activity	Phasicrate	Response peaks/ Second
Electromyographic activity	Mean of Corrugator Supercilii activity	Cor_{mean}	Micro volts
	Standard deviation of Corrugator Supercilii activity	Cor_{std}	Standard Deviation (no unit)

(*Continued*)

Table 1: (*Continued*)

Physiological Response	*Features Derived*	*Label Used*	*Unit of Measurement*
	Slope of Corrugator Supercilii activity	Cor_{slope}	Micro volts/second
	Mean IBI of blink activity	IBI $Blink_{mean}$	Milliseconds
	Standard deviation of IBI of blink activity	IBI $Blink_{std}$	Standard deviation (no unit)
	Mean amplitude of blink activity	Amp $Blink_{mean}$	Micro volts
	Standard deviation of blink activity	$Blink_{std}$	Standard deviation (no unit)
	Mean of Zygomaticus Major activity	Zyg_{mean}	Micro volts
	Std. of Zygomaticus Major activity	Zyg_{std}	Standard deviation (no unit)
	Slope of Zygomaticus Major activity	Zyg_{slope}	Micro volts/second
	Mean of Upper Trapezius activity	$Trap_{mean}$	Micro volts
	Std. of Upper Trapezius activity	$Trap_{std}$	Standard deviation (no unit)
	Slope of Upper Trapezius activity	$Trap_{slope}$	Micro volts/second
	Mean and median frequency of Corrugator, Zygomaticus, and Trapezius	$Zfreq_{mean}$ $Cfreq_{median}$ $Tfreq_{mean}$	Hertz
Temperature	Mean temperature	$Temp_{mean}$	Degree Centigrade
	Slope of temperature	$Temp_{slope}$	Degree Centigrade/ second
	Standard deviation of temperature	$Temp_{std}$	Standard deviation (no unit)

(less than 0.04 Hz). Interbeat Interval (IBI) is the time interval in milliseconds between two "R" waves in the ECG waveform. "IBI ECG_{mean}" and "IBI ECG_{std}" are the mean and standard deviation of the IBI, respectively. The R-peak detection algorithm first performed band-pass filtering on the raw ECG signal. The resulting signal was then smoothed by a 10 msec moving average window. Peaks were then detected in the resulting signal, and heuristic detection rules were applied to avoid missing R peaks or detecting multiple peaks for a single heart beat. These rules included obtaining the amplitude threshold (the difference between a peak and the corresponding inflection point) at which a peak should be considered a beat, enforcing a minimum interval of 300ms and maximum interval of 1,500 msec between peaks, checking for both positive and negative slopes in a peak to ensure that baseline drift is not misclassified as a peak, and allowing backtracking with reexamination/interpolation when peak missing was detected.

Photoplethysmograph signal (PPG) measures changes in the volume of blood in the fingertip associated with the pulse cycle, and it provides an index of the relative constriction versus dilation of the blood vessels in the periphery. Pulse transit time is the time it takes for the pulse pressure wave to travel from the heart to the periphery. Pulse transit time is determined by computing the time between systole at the heart (as indicated by the R-wave of the ECG) and the peak of the pulse wave reaching the peripheral site where PPG is being measured. Bioelectrical impedance analysis measures the impedance or opposition to the flow of an electric current through the body fluids. A common variable in recent psychophysiology research, preejection period measures the latency between the onset of electromechanical systole and the onset of left-ventricular ejection. Preejection period is derived from impedance cardiogram (ICG) time-derivative and ECG and is most heavily influenced by sympathetic innervations of the heart. The peak detection mechanisms to determine the peaks of BVP and impedance cardiogram time-derivative were similar to the ECG R-peak detection algorithm, while additional heuristic rules were added to reduce the degradation of the signal quality due to motion artifacts and avoid spurious peak detection with backtracking. Unlike ECG signals, the peak amplitudes of PPG and impedance cardiogram showed a larger deviation over a given period of time. An adaptive thresholding rule was integrated in the peak detection algorithm to address this deviation, which continuously changed/updated the threshold value to determine whether candidates for peaks qualified to be valid peaks.

Electrodermal activity consists of two main components—tonic response and phasic response. Tonic skin conductance refers to the ongoing or the baseline level of skin conductance in the absence of any particular discrete environmental events. Phasic skin conductance refers to the event related changes that occur, caused by a momentary increase in skin conductance (resembling a peak). The raw electrodermal activity signal was smoothed by a 25-msec moving average window and then down-sampled by 10 to remove the high frequency measurement noise. The phasic skin conductance detection algorithm used the following heuristics for considering a particular peak as a valid skin conductance response: (a) the slope of the rise to the peak should be greater than 0.05 microsiemens/minute, (b) the amplitude should be greater than 0.05 microsiemens, and (c) the rise time should be greater than 0.25 sec. All the signal points that were not included in the response constituted the tonic part of the skin conductance signal.

The EMG signal from Corrugator Supercilii muscle (eyebrow) captures a person's frowns and detects the tension in that region. This EMG signal is also a valuable source of blink information and helps determine the blink rate. The EMG signal from the Zygomaticus Major muscle captures the muscle movements while smiling. Upper Trapezius muscle activity measures the tension in the shoulders, one of the most common sites in the body for developing stress. The analysis of the EMG activities in the frequency domain involved applying Fast Fourier transform on a given EMG signal, integrating the EMG spectrum, and normalizing it to [0,1] to calculate the two features of interest—the median frequency and mean frequency for each EMG signal. The blink-related features were determined from the corrugator supercilii EMG signals after being preprocessed by a low-pass filter (10 Hz).

The heart sound signal measures sound generated during each heartbeat. These sounds are produced by blood turbulence primarily because of the closing of the valves within the heart. The features extracted from the heart sound signal consist of the mean and standard deviation of the third, fourth, and fifth level coefficients of the Daubechies wavelet transform. Variations in the peripheral temperature mainly come from localized changes in blood flow caused by vascular resistance or arterial blood pressure and reflect the autonomic nervous system activity. The signal was down-sampled by 10 and filtered to remove high-frequency noise, from which the time-domain features (e.g., mean, standard deviation, and slope) were calculated.

Any feature (derived from physiological signals) with an absolute correlation greater than equal to 0.3 with the target affective state was considered significant and was selected as inputs of the recognizers. It should be noted that the phenomenon of person-stereotypy makes it difficult to obtain universal patterns of emotions across individuals. As mentioned above, to overcome person-stereotypy we adopted an individual-specific approach where an affective model for each individual was developed in the Phase I study (e.g., we determine the physiological pattern of anxiety for each participant).

4. ANXIETY RECOGNITION BASED ON REGRESSION TREE

Determining a person's probable anxiety level from his or her physiological response resembles a classification problem where the attributes are physiological features and the target function is the anxiety level. The main challenge for this classification system to work, however, was the complex nature of the input physiological data sets. This complexity was primarily because of (a) high dimensionality of the input feature space (there were 46 features), (b) mixture of data types, and (c) nonstandard data structures. In addition, a few physiological data were noisy where the biofeedback sensors had picked up movement artifacts.

In our earlier work (Rani et al., 2006), we compared several machine learning algorithms, namely, K nearest neighbors (KNN), Bayesian Network technique (BNT), Support Vector Machines (SVM) and Regression Tree (RT), for affect recognition from physiological signals and found that regression tree technique was efficient for affective modeling in terms of predictive accuracy and time and space cost. Hence in this work we employed RT to determine the underlying target affective state of a player given a set of physiological features. We omit the details of the theory and learning method of RT in this article, which can be found in Breiman, Friedman, Olshen, and Stone (1984) and our previous work (Rani et al., 2006).

5. EXPERIMENTAL INVESTIGATION

5.1. Participants

Fifteen individuals (8 female, 7 male) volunteered to participate in the Phase I experiments. Their age ranged from 18 to 34 years, except for one participant,

who was 54 years old. They were from diverse professional and ethnic backgrounds. They all had college degrees and had experience of playing computer games. Because of the nature of the tasks, the following were considered when choosing the participants: (a) their fluency in English, (b) their familiarity with computers and ease of operation of keyboard and mouse, and (c) general health (the absence of any problem in hearing or sensing). Participants were solicited through phone, emails and flyers posted around the Vanderbilt University area. They were given monetary compensation for their voluntary participation. Out of the 15 participants, 9 also took part in Phase II experiments.

The Institutional Review Board approval was sought and received for conducting these experiments. In the Institutional Review Board application, all details of the experiment were reported and it was emphasized that the health and safety of the participants was by no means endangered by participating in these experiments. It was also mentioned that the range of anxiety that the participants could experience was no more intense than the levels of anxiety that are commonly experienced in daily life. A detailed consent form was also drafted that acquainted the participants with the experimental procedure and their role in it. Participants were allowed to participate in the experiment only after their consent had been obtained through a signed consent form.

5.2. Game Design

Two computer games were designed and implemented to evoke varying intensities of anxiety from the participants. Physiological data from participants were collected during the experiments. The two games consisted of solving anagram and playing Pong. The anagram game has been previously employed to explore relationships between both electrodermal and cardiovascular activity with anxiety (Pecchinenda & Smith, 1996). Emotional responses were manipulated in this game by presenting the participants with anagrams of varying difficulty levels, as established through pilot work. An optimal mix of solvable anagrams caused low level of anxiety at times. Unsolvable or extremely difficult anagrams and giving time deadlines generated anxiety. In Pong sessions the participants played a variant of the classic computer game, "Pong." This game has also been used in the past by researchers to study anxiety and performance (Brown et al., 1997). Various parameters of the game were manipulated to elicit anxiety. These included ball speed and size, paddle speed and size, sluggish or overresponsive keyboard, and random keyboard response. The anxiety levels ranged from a low level of anxiety caused by a low ball speed and large sizes of the ball and the paddle, to high level of anxiety caused by a very high ball speeds and sluggish or overresponsive keyboard.

Each game session was subdivided into a series of discrete epochs that were bounded by self-reported affective state assessments. During the assessment, participants reported their perceived level of anxiety on a pop-up dialog box. It occurred every 3 min for the anagram game and every 2 to 4 min for the Pong game. This information was collected using a battery of self-report questions rated on a 9-point Likert scale, where 1 indicated the lowest level and 9 indicated

the maximum level. The reported level of anxiety were labeled and used for affective modeling in Phase I and assessing the real-time prediction performance of affective model in Phase II.

Based on a previous pilot study, different configurations of game parameters were determined to vary the difficulty level. During piloting, participants played a number of epochs of each game with selected configurations. After each epoch, the difficulty of the configuration perceived by them (on a 9-point Likert scale) were reported and recorded, as well as their performances, such as the number of correct answers for anagram game and the number of balls that they successfully hit in Pong game. After the piloting was over, these results were compiled to determine the perceived difficulty level of each configuration. The configurations were sorted and grouped according to their difficulty ratings. It was found that there were three distinct clusters of configurations that were well separated along the difficulty scale. These clusters were named Levels I, II and III, in the increasing order of difficulty. The averaged performance of participants for a given configuration was used to determine the threshold for that configuration.

In Phase II study, Pong game was used to study the impacts of an affect-based mechanism and a performance-based DDA mechanism to the gaming experience. The target number of hits (*TNH*) was defined as 10% higher than the average across the thresholds of all configurations for a given difficulty level. After each epoch was over, the participant's performance was rated as excellent ($hits \geq \lfloor 1.2TNH \rfloor$), good ($\lfloor 0.8TNH \rfloor \leq \text{hits} < \lfloor 1.2TNH \rfloor$) or poor ($hits < \lfloor 0.8TNH \rfloor$).

5.3. Phase I: Affective Modeling

To obtain physiological data to build affective models, the experiment in Phase I were designed to elicit varying intensities of the target affective state in participants as they played the computer games. We only provide brief but necessary information regarding affective modeling here. A detailed description of Phase I work can be found in (Rani et al., 2006). The training data set consisted of labeled self-reports of anxiety and various features (as described in section 3) that were extracted off-line from the collected physiological data. The affective model was developed using the Regression Tree method. Each participant took part in six sessions—three anagram game sessions and three Pong game sessions. Each session was approximately 1 hr long and consisted of 16 epochs on an average. An epoch was a 2 to 4 min followed by self-reporting that usually lasted for an interval of 30 sec to 1 min. After the self-reporting the next epoch would begin. At the beginning of each session, baseline physiological signals were recorded in order to offset day-variability. Phase I study spanned a period of about two months.

To develop affective model, we built mappings from the extracted physiological features to the intensity (i.e., low/medium/high) of anxiety. This mapping was cast as a classification problem. The training datasets were formed by merging physiological features and self-reports of the participants as shown in Figure 1. The physiological data and self-reports were recorded in two separate files at the time of experiment. Later, the physiology data file was partitioned into the data blocks pertaining to every epoch in a separate file. Then each epoch file was

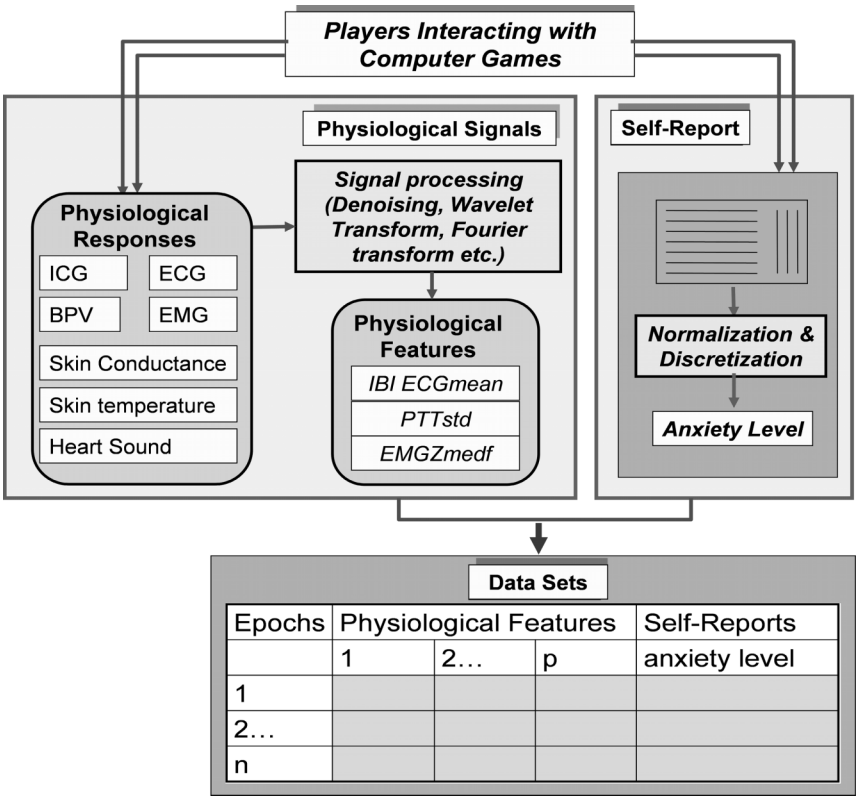

FIGURE 1 Formation of data sets for affective modeling.

processed to extract the relevant features from the physiology data. The reports on anxiety was normalized to [0, 1] and then discretized such that 0–0.33 was labeled low, 0.34–0.67 was medium, and 0.68–1.0 was labeled high. During the experiment, 15 data sets were collected (1 for each participant). Each data set has n ($n \approx 96$) data vectors. The prediction accuracy of the developed model was evaluated by the leave-one-out cross validation method. Results of affective modeling can be found in Section 6.1.

5.4. Phase II: Affect-Based Dynamic Difficulty Adjustment

In Phase II, two sessions of the Pong game with two different DDA mechanisms were conducted for each participant: one in which the game difficulty was adapted based on player's performance; and another in which the real-time recognized player's anxiety level was employed to alter game difficulty. Phase II study spanned a period of about 1 month.

Experiment setup. The setup for the Pong game, which can adjust difficulty level dynamically based on recognized affective state/performance, is shown in

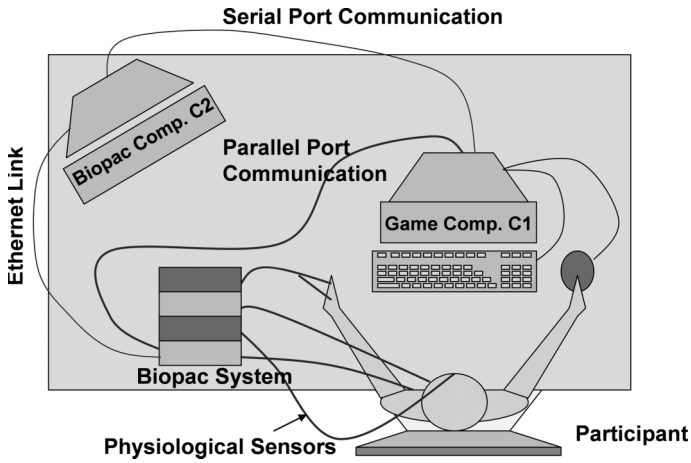

FIGURE 2 Phase II experimental setup.

Figure 2. The participant played the game on computer C1 while his or her physiological data were acquired via the Biopac system connected to C2. Physiological signals were transmitted from the Biopac transducers to C2 through an Ethernet link at 1,000 Hz after being amplified and digitized by the Biopac system. C1 was also connected to the Biopac system via a parallel port, through which the game related event markers were recorded along with the physiological data in a time-synchronized manner. The serial communication between C1 and C2 enabled them to communicate with each other. C2 performed the following functions: (a) established serial communication with C1 to acquire the performance rating of the participant, (b) acquired signals from Biopac system (that included the physiological signals and the event markers), (c) ran signal processing routines to process the physiological data to extract features online, and (d) used affective model developed in Phase I to recognize the anxiety level in real time. Hence C2 could determine the affective state of the participant as well as his or her current performance. This information along with the knowledge of the current game difficulty was utilized to determine the next level of difficulty of the game. There was a serial communication protocol established between C1 and C2 that ensured that begin/end of Pong epochs on C1 was appropriately synchronized with the physiological data acquisition on C2.

Experimental design. Nine of the 15 participants who had taken part in Phase I study participated in Phase II experiments. Each of these 9 participants played two Pong-playing sessions (Png1 and Png2). In Png1, the game difficulty was adjusted based on performance without any regard to the anxiety of the participant, whereas In Png2, the game difficulty was changed based on the real-time recognized anxiety level of the participant without regard to the performance.

Each Pong session was approximately 45 min and consisted of 12 epochs of 2 min each. The remaining time was spent in setup, attaching sensors, getting self-reports, and taking breaks. After every epoch, the participant reported his or her

assessment of one's own anxiety on a 9-point Likert scale. In addition, at the end of each of the whole completed session, participants answered questions pertaining to their overall experience during the entire session on a 9-point Likert scale, which included overall anxiety, enjoyment, challenge, and self-evaluation of the performance. These questions were asked to determine the aggregate gaming experience of each completed session (as described in section 6.2).

During any Pong epoch, the game proceeded as follows:

1. A pop-up dialog box describing the rules of the game and other game-related instructions appeared on the game computer.
2. The participant was notified of the goal (number of minimum hits, maximum allowable misses and the time available) via a pop-up dialog box on the game computer.
3. Once the game started, the participant used the up and down arrow keys on the computer to control the paddle to hit the moving ball on-screen.
4. During any given epoch, the number of hits, misses, and seconds remaining were continuously updated on the bottom panel of the Pong screen.
5. After each epoch was over, the participant's performance was rated as excellent, good, or poor.
6. The end of a given epoch was followed by an interval of 30 sec to 1 min for self-reporting. After the self-reporting was completed, the next epoch would begin.

In Pong game in Phase II, three levels of difficulty—Level I (easy), Level II (moderately difficult), and Level III (very difficult)—and three levels of performance—poor, good, and excellent—were identified. We also classified anxiety in three levels—low, medium, and high. Figures 3 and 4 show the state-flow models that were utilized to dynamically adjust difficulty level based on performance and anxiety, respectively. It can be seen that in the performance-based DDA,

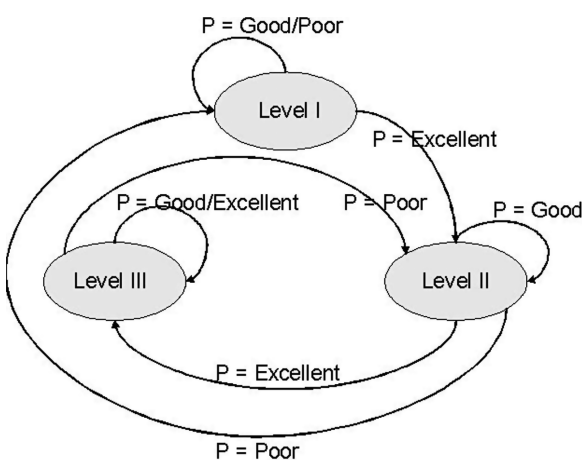

FIGURE 3 State-flow diagram for performance-based dynamic difficulty adjustment.

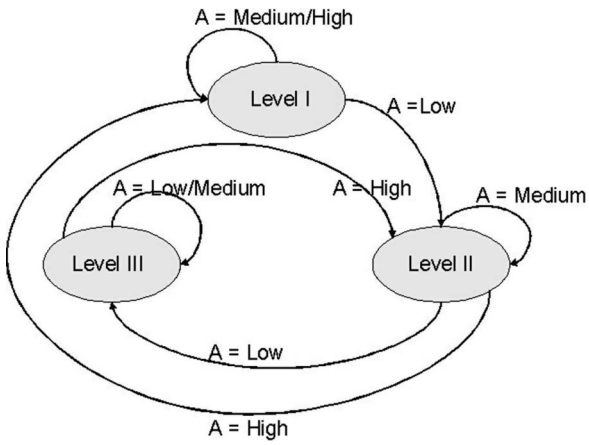

FIGURE 4 State-flow diagram for affect-based dynamic difficulty adjustment.

excellent performance resulted in an increase in the level of difficulty (except when the player was already at the highest level), good performance caused the level to remain constant at the current level, and poor performance resulted in a decrease in difficulty level (except when the player was already at the lowest level). In the affect-based DDA, it can be seen that low anxiety resulted in increase in the level of difficulty (except when the player was already at the highest level), medium anxiety caused the level to remain constant at the current level, and high anxiety resulted in a decrease in difficulty level (except when the player was already at the lowest level).

The following conditions were imposed to avoid bias in data due to habituation, session-order, and to deal with day-variability: (a) to prevent habituation, an interval of at least 10 days between any two Pong sessions was enforced; (b) the sessions (performance based and affect based) were randomized to avoid any bias due to the order of sessions; and (c) all the other experimental conditions were kept constant over all sessions.

6. RESULTS AND DISCUSSION

In this section we first briefly discuss the Phase I result that presents the off-line performance of the affective models. We then discuss the Phase II results from real-time closed-loop experiments in detail, which are the primary contribution of this work.

6.1. Phase I: Off-Line Affective Modeling

In Phase I, we performed a comparative assessment of several machine learning methods for developing affective models. Figure 5 shows that the mean percentage accuracy (averaged across all the game epochs for all participants) to distinguish

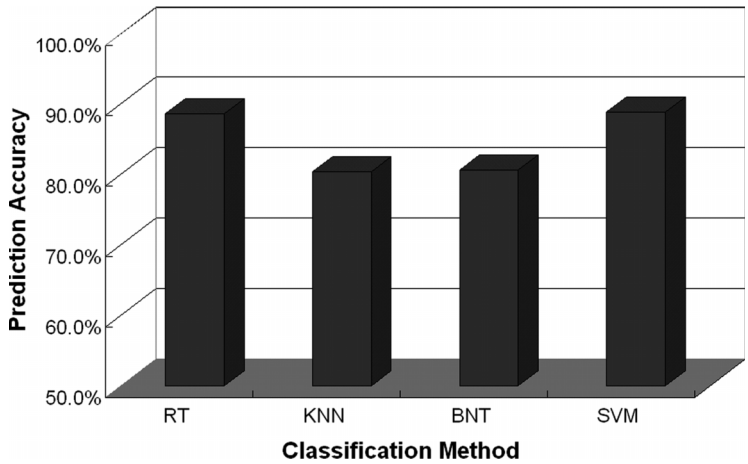

FIGURE 5 Prediction accuracy for affective state of anxiety. *Note:* RT = Regression Tree; KNN = K Nearest Neighbors; BNT = Bayesian Network Technique; SVM = Support Vector Machines.

between different levels of anxiety were 88.5% for RT, 80.4% for KNN, 80.6% for BNT, and 88.9% for SVM. The results showed that all the above-mentioned methods performed well. This was in accordance with the claim of psychophysiologists that there is a distinct relationship between physiology and underlying affective states. Among the four machine learning techniques that were examined, both RT and SVM gave more reliable classification accuracy. However, it should be noted that RT does not require any parameter tuning, whereas in the case of SVM, choosing appropriate parameters (e.g., regularization parameter and kernel parameters) was imperative (Vapnik, 1998). A detailed analysis of the time and space efficiency of RT-based affective modeling can be found in our previous work (Rani et al., 2006). Because regression tree technique was efficient for anxiety recognition, it was used in Phase II for game difficulty adjustment.

6.2. Phase II: Comparison of Affect-Based DDA with Performance-Based DDA in Real-Time Gaming

The results presented here are based on the validation Pong game sessions: Png1 (with performance-based DDA) and Png2 (with affect-based DDA). We observed several important results, summarized next:

- *The real-time prediction accuracy of the affective models was high.* Once the affective modeling is accomplished in Phase I, the model can accept as input the physiological features, extracted online, and produce as output the probable level of anxiety of a participant when he/she is playing the computer game. The average real-time prediction accuracy, which represents how closely the

online physiology-based quantitative measure of anxiety level matched with that of the subjective rating of anxiety, was 78% across all the 9 participants. Note that our affective model was evaluated based on physiological data obtained online from a real-time application. However, even then this real-time prediction accuracy is comparable to the results achieved through off-line analysis as reported in the literature (Kim et al., 2004; Nasoz et al., 2003; Rani, 2005).

- *The performance of the majority of the participants improved during the affect-based DDA session.* The improvement in performance after the performance-based and anxiety-based sessions was shown in the "Performance" column of Table 2. In each of these sessions, the first and the last epoch were identical test epochs and the difference in the number of hits of the last and the first epoch gave the performance improvement. As can be seen that 7 out of 9 participants showed a greater improvement in performance after the affect-based session while 2 did not show any improvement (participants 5 and 9). Among the 7 participants who showed an improvement in affect-based game adaptation, 2 participants actually had a degradation of performance during the performance-based game adaptation. Using repeated measure analysis of variance (ANOVA) test, it was observed that the null hypothesis (asserting that there was no change in performance between performance-based and anxiety-based game sessions) could be rejected ($p < .05$).
- *Most participants perceived the game with the affect-based DDA to be more challenging than the one with the performance-based DDA.* At the end of each completed game session, the participants had reported the level of challenge that they had experienced and from this self-report, it was seen that most participants perceived the game with the affect-based DDA to be more challenging than the one with the performance-based DDA (Challenge column in Table 2). Except P1 and P5 who reported constant challenges across the two sessions, all the other participants reported an increase in challenge during the anxiety-based session. Using repeated measure ANOVA test, it was observed that the null hypothesis (asserting that there

Table 2: Perceived Anxiety, Performance Improvement, Challenge, and Satisfaction Index (SI) Across Performance-Based (P) and Anxiety-Based (A) Sessions

	Performance		*Challenge*		*SI*		*Anxiety*	
Participant ID	*P*	*A*	*P*	*A*	*P*	*A*	*P*	*A*
P1	2	5	5	5	13	13	5	1
P2	5	10	5	7	13	16	6	4
P3	1	5	5	7	10	17	6	3
P4	–3	10	4	5	14	11	2	6
P5	10	10	7	7	15	19	5	7
P6	3	6	8	9	22	23	8	5
P7	22	24	5	7	18	21	8	6
P8	–6	12	4	5	13	18	7	5
P9	0	0	4	7	14	16	5	5

was no change in challenge between performance-based and anxiety-based sessions) could be rejected ($p < .01$).

- *Most participant perceived that the game with the affect-based DDA to be more satisfying than the one with the performance-based DDA.* An index called Satisfaction Index (*SI*) was defined by combining the values of challenge (*C*), enjoyment (*E*), and performance appraisal (*P*) reported by the participants at the end of each session.

$$SI = C + E + P \tag{1}$$

The *SI* could be a possible measure of the overall satisfaction of the participant during a given game session controlled by either the performance-based DDA or the affect-based DDA. There have been many efforts to develop metrics for measuring enjoyment in computer games, but no formal standards have yet been developed for evaluating fun, enjoyment, or satisfaction. Echoing similar opinion, Wiberg (2005) stated, "Research into the aspect of user satisfaction has so far been neglected in the research discipline of HCI. . . . When discussing fun and entertainment in the context of usability, the most closely related notion is 'user satisfaction.'" In the work by Sweetser and Wyeth (2005), the authors presented a model of enjoyment based on eight elements—concentration, challenge, skills, control, clear goals, feedback, immersion, and social interaction. They claimed that each of these elements contributes to achieving enjoyment in games. We used challenge and skill (as indicated by performance) along with a direct report on enjoyment to compute the *SI*.

The SI column of Table 2 shows the values of *SI* during the two game sessions for each participant. Seven of 9 participants reported an increase in the *SI* during the game session with the affect-based DDA. Of the 2 participants who did not report higher satisfaction during the anxiety-based session, P1 reported no change where as P4 reported a decrease in overall satisfaction. It should be noted that P4 also reported an increase in anxiety during the affect-based session, and P1 reported no change in perceived challenge during the two sessions. Using repeated measure ANOVA test, it was observed that the null hypothesis could be rejected ($p < .05$).

- *The perceived anxiety-level was reduced for the majority of the participants during the affect-based DDA session*. The results discussed so far suggest that anxiety-based DDA has positively influenced user satisfaction, feeling of challenge, and performance. In addition, we were also interested to know how the participant felt about their anxiety during gaming. The anxiety of the participants as reported by them (perceived anxiety) at the end of the completed affect-based session and the completed performance-based session was shown in the Anxiety column of Table 2. It can be seen that out of 9 participants, 6 reported a decrease in anxiety, 2 reported an increase, and 1 reported no change in anxiety during the anxiety-based session as compared to the performance-based session. Although the majority of the participants felt that they were less anxious when playing the game with affect-based

DDA, it is interesting to note that no statistically significant difference in perceived anxiety was observed between the affect-based sessions and the performance-based sessions (p = .24, repeated measure ANOVA) when using the reports of perceived anxiety collected at the end of the session. To explore the nature of perceived anxiety during the gaming process we analyzed the anxiety reports after each epoch that we believe represent a more accurate record of perceived anxiety over the entire game duration. As described in section 5.4.2, each session consisted of 12 epochs and besides the reports at the end of each completed session, a participant also reported his or her assessment of one's own anxiety after every epoch. This epoch-based anxiety reports may allow a finer grain analysis on the difference of perceived anxiety during the process of the game playing in the two conditions (anxiety based vs. performance based). A nested random-effect mixed model test was performed to evaluate the significance of the anxiety difference between the two sessions, and it was observed that such difference was statistically significant (M = 3.41 for affect-based sessions and M = 4.42 for performance-based session, $p < .01$). Given the facts that majority of the participants felt less anxious after the affect-based sessions and that there was significant differences in the perceived anxiety during the game playing process in the two conditions, it suggests that by utilizing the information regarding the probable anxiety level of the participant to continuously adapt the game difficulty, the affect-based DDA has the potential to impact the gaming experience positively and keep the participants in a lower anxiety state.

7. CONCLUSIONS

In recent years several researchers have investigated DDA mechanisms to improve game-playing experiences such that the games can be automatically tailored to individual characteristics. However, most existing works on DDA mechanisms focus on player's performance as the determining factor. These DDAs do not possess the ability of deciphering affective cues of the players. Although performance assessment is important and useful, affective states of the players can have major impacts on the gaming experience. This paper reported our efforts in developing an affect-based DDA mechanism to allow a computer game to infer and respond to the affective state while interacting with the players. The affective state (e.g., anxiety, in this case) was recognized using psychophysiological analysis. We explored a comprehensive set of physiological indices for affective modeling. The gaming experiences of the participants were evaluated and compared when a performance-based DDA mechanism and an affect-based DDA mechanism were applied to the same computer game. This is the first time, to our knowledge, that the impacts of an affect-base DDA to player's interaction with a computer game that is capable of physiology-based affect recognition and real-time difficulty adjustment in a closed-loop manner has been investigated experimentally.

Four machine learning methods were investigated to classify the anxiety level. A Regression Tree based affective model yielded reliable prediction with approximately

88% success, whereas the other three approaches also performed competitively. When the developed RT-based model was applied in Phase II to recognize the anxiety level during the game play in real time, it gave 78% correct predictions. Although relatively less existing works investigated affect recognition in real-time applications and although further exploration in this direction is needed, this result suggested that physiology-based affective modeling provides a promising methodology to objectively quantify player's emotion when interacting with computer games. A systematic experimental study was conducted to evaluate the impacts of an affect-based DDA on the game play by comparing it with a performance-based DDA. It was observed that 6 of 9 participants showed lower anxiety during the anxiety-based session than in the performance-based session, and 7 participants showed a greater improvement in performance during the anxiety-based session. Of the participants, 77% reported more challenging gaming experience and the overall satisfaction of gaming was enhanced by the affect-based DDA for majority of participants. These results suggest that gaming experience could be enhanced when a computer game is capable of recognizing player's affective states and adjusting game difficulty accordingly.

Note that the presented work requires physiological sensing that has its own limitations. For example, one needs to wear physiological sensors, and use of such sensors could be restrictive under certain circumstances. However, given the rapid progress in wearable computing, for example, physiological sensing clothing and accessories (Jafari, Dabiri, Brisk, & Sarrafzadeh, 2005; Sung & Pentland, 2005; Wijesiriwardana, Mitcham, & Dias, 2004), we believe that physiology-based affect recognition can be appropriate and useful to achieve affect-sensitive gaming.

One limitation of this work is that six 1-hr gaming sessions were conducted for each participant to collect the training data for affective modeling. Further work is needed to reduce the length of time and the data for model building so that the affect-based DDA can be efficiently applied to game applications. The next research goal would be to explore the trade-off between prediction accuracy and training set size and investigate new machine learning techniques to optimize training data to compensate for its scarcity. Active learning (Vijayakumar & Ogawa, 1999) is one method that could hold promise for such a purpose. Active learning method can assume some control over what next game epoch to be introduced during the affective modeling process to get a more informative training point. It is also expected that the required training process would be reduced when the player's physiology is used together with other channels of affect-related information, such as eye gaze (Prendinger, Ma, & Ishizuka, 2007) and posture (Tan, Slivovsky, & Pentland, 2001). The presented work, however, demonstrated the feasibility that a player's affective state can be deciphered from his/her physiological response during gaming and a DDA mechanism can be designed that can adjust the game difficulty in real-time based on the affective state information. The experimental investigation showed the benefits of such a DDA mechanism. It is expected these results will encourage future research into affect-based DDA design for computer games. In addition, besides anxiety, other affective states (e.g., excitement and frustration) are also considered to be important in game playing (Gilleade & Dix, 2004; Mandryk & Atkins, 2007). Although the affective modeling methodology in this work could be used to detect the intensity

of anxiety, excitement, and frustration simultaneously, more sophisticated difficulty adaptation mechanisms would be demanded to incorporate multiple inferred affective cues and account for other game playing information of interests, such as the player's performance, his or her personality, and the context and complexity of the game. We will investigate fast and robust DDA mechanisms that would permit a computer's adaptive response in the more complex gaming applications and allow the affect-sensitive DDA to be adopted in the future computer games.

REFERENCES

Andrade, G., Ramalho, G., Santana, H., & Corruble, V. (2005). *Challenge-sensitive action selection: An application to game balancing.* Paper presented at the IEEE/WIC/ACM International Conference on Intelligent Agent Technology Compiègne, France.

Asha, K., Ajay, K., Naznin, V., George, T., & Peter, F. D. (2005). *Gesture-based affective computing on motion capmture data.* Paper presented at the Int. Conf. on Affective Computing and Intelligent Interaction (ACII), Beijing, China.

Bartlett, M. S., Littlewort, G., Fasel, I., & Movellan, J. R. (2003). *Real time face detection and facial expression recognition: development and applications to human computer interaction*. Paper presented at the Computer Vision and Pattern Recognition Workshop, June, Madison, Wisconsin.

Bloom, J., & Trautt, G. M. (1978). Finger pulse volume as a measure of anxiety: Further evaluation. *Psychophysiology*, *14*(6), 541–544.

Bradley, M. M. (2000). Emotion and motivation. In J. T. Cacioppo, L. G. Tassinary & G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 602–642). New York: Cambridge University Press.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Boca Raton, FL: Chapman & Hall/CRC.

Brown, R. M., Hall, L. R., Holtzer, R., Brown, S. L., & Brown, N. L. (1997). Gender and video game performance. *Sex Roles*, *36*(11–12), 793–812.

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, 50(4), 31–34.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, *16*(7–8), 555–575.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., et al. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1), 32–80.

Dawson, M. E., Schell, A. M., & Filion, D. L. (1990). The electrodermal system. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Principles of psychophysiology: Physical, social, and inferential elements* (pp. 295–324). New York: Cambridge University Press.

Demasi, P., & Cruz, A. (2002). *Online coevolution for action games*. Paper presented at the 3rd International Conference on Intelligent Games and Simulation, November, London.

Ekman, P., & Friesen, W. V. (1986). A new pan cultural facial expression of emotion. *Motivation and Emotion*, *10*(2), 159–168

Gilleade, K. M., & Allanson, J. (2003). *A toolkit for exploring affective interface adaptation in videogames*. Paper presented at the HCI International 2003, June, Crete, Greece. Mahwah, New Jersey: Lawrence Erlbaum, pp. 22–27.

Gilleade, K. M., & Dix, A. (2004). *Using frustration in the design of adaptive videogames*Paper presented at the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology, June, Singapore.

Gilleade, K., Dix, A., & Allanson, J. (2005). Affective videogames and modes of affective gaming: Assist me, challenge me, emote me. *Proceedings of DIGRA 2005,* June 2005, Vancouver, Canada. pp. 16–20.

Hunicke, R., & Chapman, V. (2004). AI for dynamic difficult adjustment in games. *Challenges in Game Artificial Intelligence AAAI Workshop* (pp. 91–96). San Jose.

Jafari, R., Dabiri, F., Brisk, P., & Sarrafzadeh, M. (2005). *Adaptive and fault tolerant medical vest for life critical medical monitoring*. Paper presented at the the 20th ACM Symposium on Applied Computing, March, Santa Fe, NM.

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical & Biological Engineering & Computing*, *42*(3), 419–427.

Koster, R. (2004). *Theory of fun for game design*. Phoenix, AZ: Paraglyph Press.

Kramer, A. F., Sirevaag, E. J., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated fight missions. *Human Factors*, *29*(2), 145–160.

Kulic, D., & Croft, E. (2007). Physiological and subjective responses to articulated robot motion. *Robotica*, *25*, 13–27.

Lacey, J. I., & Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. *American Journal of Psychology*, *71*(1), 50–73.

Lee, C. M., & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*, *13*(2), 293–303.

Leon, E., Clarke, G., Callaghan, V., & Sepulveda, F. (2004). Real-time detection of emotional changes for inhabited environments. *Computers & Graphics-Uk*, *28*(5), 635–642.

Leon, E., Clarke, G., Callaghan, V., & Sepulveda, F. (2007). A user-independent real-time emotion recognition system for software agents in domestic environments. *Engineering Applications of Artificial Intelligence*, *20*(3), 337–345.

Magerkurth, C., Cheok, A. D., Mandryk, R. L., & Nilsen, T. (2005). Pervasive games: bringing computer entertainment back to the real world. *ACM Computers in Entertainment*, *3*(3), 11–29.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies*, *65*(4), 329–347.

Nasoz, F., Alvarez, K., Lisetti, C. L., & Finkelstein, N. (2003). Emotion recognition from physiological signals for presence technologies. *International Journal of Cognition, Technology, and Work—Special Issue on Presence*, *6*(1), 4–14.

Pagani, M., Lombardi, F., & Guzzetti, S. (1986). Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympathovagal interaction in man and conscious dog. *Circulation Research*, 59, 178–193.

Pagulayan, R. J., Keeker, K., Wixon, D., Romero, R. L., & Fuller, T. (2002). User-centered design in games. In J. A. Jacko & A. Sears (Eds.), *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications* (pp. 883–906). Mahwah, NJ: Erlbaum.

Pagulayan, R. J., Steury, K. R., Fulton, B., & Romero, R. L. (2005). Designing for fun: user testing case studies. In M. A. Blythe et al. (Eds), *Funology: From usability to enjoyment* (pp. 137–150). Dordrecht, the Netherlands: Kluwer Academic.

Pecchinenda, A., & Smith, C. A. (1996). The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition & Emotion*, *10*(5), 481–503.

Picard, R. W. (1997). *Affective computing*. Cambridge, MA: MIT Press.

Prendinger, H., Ma, C. L., & Ishizuka, M. (2007). Eye movements as indices for the utility of life-like interface agents: A pilot study. *Interacting With Computers*, *19*(2), 281–292.

Prendinger, H., Mori, J., & Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human–Computer Studies*, *62*(2), 231–245.

Rani, P. (2005, December). *Psychophysiology based affective communication for implicit human-robot interaction*. Unpublished doctoral dissertation,Vanderbilt University, Nashville, TN.

Rani, P., Liu, C. C., Sarkar, N., & Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis and Applications*, *9*(1), 58–69.

Rani, P., Sarkar, N., Smith, C. A., & Kirby, L. D. (2004). Anxiety detecting robotic system - towards implicit human-robot collaboration. *Robotica*, 22, 85–95.

Rani, P., Sims, J., Brackin, R., & Sarkar, N. (2002). Online stress detection using psychophysiological signals for implicit human-robot cooperation. *Robotica*, 20, 673–685.

RocResearch. (2004). *Video game industry*. London, UK: RocResearch Ltd.

Rohrmann, S., Hennig, J., & Netter, P. (1999). Changing psychobiological stress reactions by manipulating cognitive processes. *International Journal of Psychophysiology*, *33*(2), 149–161.

Sykes, J., & Brown, S. (2003). Affective gaming. Measuring emotion through the gamepad. *Proceedings of CHI-03*, 732–733.

Spronck, P., Ponsen, M., Sprinkhuizen-Kuyper, I., & Postma, E. (2006). Adaptive Game AI with Dynamic Scripting. *Machine Learning*, *63*(3), 217–248.

Stokes, B. (2005). Videogames have changed: Time to consider serious games? *The Development Education Journal*, *11*(3), 12–14.

Sung, M., & Pentland, A. S. (2005). Minimally-Invasive physiological sensing for human-aware interfaces. *HCI International 2005*. Las Vegas, NV.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *Computers in Entertainment*, *3*(3), 3–3.

Tan, H. Z., Slivovsky, L. A., & Pentland, A. (2001). A sensing chair using pressure distribution sensors. *IEEE-ASME Transactions on Mechatronics*, *6*(3), 261–268.

Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.

Vicente, K. J., Thornton, D. C., & Moray, N. (1987). Spectral-analysis of sinus arrhythmia—A measure of mental effort. *Human Factors*, *29*(2), 171–182.

Vijayakumar, S., & Ogawa, H. (1999). Improving generalization ability through active learning. *IEICE Transactions on Information and Systems*, E82d(2), 480–487.

Watts, J. M. J. (1975). Anxiety and the habituation of the skin conductance response. *Psychophysiology*, *12*(5), 596–601.

Wiberg, C. (2005). Affective computing vs. usability? Insights of using traditional usability evaluation methods, *CHI 2005 Workshop on Innovative Approaches to Evaluating Affective Interfaces*.

Wijesiriwardana, R., Mitcham, K., & Dias, T. (2004). *Fibre-meshed transducers based real time wearable physiological information monitoring system*. Paper presented at the International Symposium on Wearable Computers, Washington, DC.

Zaphiris, P. (2007). HCI issues in computer games. *Interacting With Computers*, *19*(2), 135–139.