Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 65 (2007) 329–347
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies

Regan L. Mandryk*, M. Stella Atkins

School of Computing Science, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

Available online 10 January 2007

Abstract

The popularity of computer games has exploded in recent years, yet methods of evaluating user emotional state during play experiences lag far behind. There are few methods of assessing emotional state, and even fewer methods of quantifying emotion during play. This paper presents a novel method for continuously modeling emotion using physiological data. A fuzzy logic model transformed four physiological signals into arousal and valence. A second fuzzy logic model transformed arousal and valence into five emotional states relevant to computer game play: boredom, challenge, excitement, frustration, and fun. Modeled emotions compared favorably with a manual approach, and the means were also evaluated with subjective self-reports, exhibiting the same trends as reported emotions for fun, boredom, and excitement. This approach provides a method for quantifying emotional states continuously during a play experience.

Keywords: Emotion; Affective computing; Fun; Games; Physiology; Galvanic skin response; Electromyography; Heart rate

1. Introduction

Computer games have grown during recent years into a very popular entertainment form with a wide variety of game types and a large consumer group spread across the world. As researchers develop novel play environments, computer and console game markets continue to grow rapidly, outperforming the film industry in terms of total revenues in many regions (Pagulayan et al., 2002). Although gaming technology has continued to evolve, researchers and traditional computer game developers suffer from a lack of effective evaluation methods.

The development of evaluation methodologies in human–computer interaction research (HCI) has been rooted in the cognitive sciences of psychology and human factors, in the applied sciences of engineering, and in computer science (Norman, 2002). Although the study of human cognition has made significant progress in the last decade, the idea of emotion, which is equally important to design (Norman, 2002), is still not well understood, especially when the primary goals are to challenge and entertain the user. Traditional measures for productivity environments, such as task performance, are not applicable to affective environments since we are not interested in performance; we are interested in what kind of emotional experience is provided by the play technology and environment, regardless of performance (Pagulayan et al., 2002). Although traditional usability measures may still be relevant, they are subordinate to the emotional experiences resulting from interaction with the play technology and with other players.

Our research interest is in how to quantify emotional experience when engaged with affective technologies, by developing an evaluation methodology for entertainment environments that is as robust as methods for evaluating productivity. This paper motivates why we need such an approach and describes the process by which we designed a new evaluative methodology for measuring emotional experience with interactive entertainment technologies.

*Corresponding author. Faculty of Computer Science, Dalhousie University, 6050 University Ave., Halifax, NS, B3H 1W5, Canada. Tel.: +1 902 494 1040; fax: +1 902 492 1517.

E-mail addresses: rlmandry@cs.sfu.ca (R.L. Mandryk), stella@cs.sfu.ca (M.S. Atkins).

1071-5819/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2006.11.011

1.1. Affective evaluation of entertainment technologies

Traditional evaluation methods have been adopted, with some success for evaluating entertainment technologies, and include both subjective and objective techniques. The most common methods of assessing emotion are through subjective self-reports including questionnaires, interviews, and focus groups (Fulton and Medlock, 2003) and through objective reports from observational video analysis (Lazzaro, 2004).

The success of a play environment is determined by the *process* of playing, not the *outcome* of playing (Pagulayan et al., 2002). We must consider this when evaluating emotional experience during interaction with play technologies, as current methods suffer from low evaluative bandwidth, providing information on the whole experience, rather than continuously throughout time.

Subjective reporting through questionnaires and interviews is generalizable, and is a good approach to understanding the *attitudes* of the users, but subjects are bad at self-reporting their *behaviors* in game situations (Pagulayan et al., 2002). In addition, subjective techniques only generate data when a question is asked, and interrupting game play to ask a question is too disruptive. Desmet (2003) developed a non-verbal questionnaire designed specifically to assess 14 separate emotional responses to products. Although it addresses some of the drawbacks of language scales, the evaluative bandwidth is still low.

Using video to code gestures, body language, facial expressions and verbalizations, is a rich source of data; however, there is an enormous time commitment, which requires between 5 and 100 h of analysis for every hour of video (Fisher and Sanderson, 1996).[1] Also, the analysis is generally event-based (user is smiling now), rather than continuous (degree of smile for every point in time), which could be important for exploring the process of play.

There has been some recent research on using inspection methods, such as heuristics (Wiberg, 2003; Desurvire et al., 2004; Sweetsner and Wyeth, 2005) to evaluate the playability of an entertainment technology, but these discount methods do not involve actual users, but are administered by usability specialists. Heuristics also give an overview of the playability, rather than examining a user's change in emotions over time.

Researchers in human factors have used physiological measures as indicators of mental effort and stress (Vicente et al., 1987). See Mandryk and Inkpen (2004) for an overview. Psychologists use physiological measures to differentiate human emotions such as anger, grief, and sadness (Ekman et al., 1983). Recently, physiological measures have been used to assess a user's emotional experience when engaged with computing systems (see Section 2.4); however, physiological data have not been employed to identify a user's emotional state, such as fun or excitement, when engaged with play technologies. Based on previous research on the use of psychophysiological techniques, we believe that capturing, measuring, and analyzing autonomic nervous system (ANS) activity will provide researchers and developers of technological systems with continuous access to the emotional experience of the user. Used in concert with other evaluation methods (e.g. subject reports and video analysis), a complex, detailed account of both conscious and subconscious user experience could be formed.

We designed an experiment to create and evaluate a model of user emotional state when interacting with play technologies. We record users' physiological, verbal and facial reactions to game technology, and apply post-processing techniques to quantitatively and continuously measure emotional state. We envision that when combined with other evaluative approaches, our technique can help create a rich and robust picture of user experience.

2. Physiological metrics for evaluation

In this section we briefly introduce the physiological measures used, describe how these measures are collected, and explain their inferred meaning. Based on previous literature, we chose to collect galvanic skin response (GSR), electrocardiography (EKG), and electromyography of the face ($EMG_{smiling}$ and $EMG_{frowning}$). Heart rate (HR) was computed from the EKG signal.

2.1. Galvanic skin response

GSR is a measure of the conductivity of the skin. There are specific sweat glands (eccrine glands) that cause skin conductivity to change and result in the GSR. Located in the palms of the hands and soles of the feet, these sweat glands respond to psychological stimulation rather than simply to temperature changes in the body (Stern et al., 2001). For example, many people have cold clammy hands when they are nervous. In fact, subjects do not have to even be sweating on the palms of the hands or soles of the feet to see differences in GSR because the eccrine sweat glands act as variable resistors on the surface. As sweat rises in a particular gland, the resistance of that gland decreases even though the sweat may not reach the surface of the skin (Stern et al., 2001).

Galvanic skin response is a linear correlate to arousal (Lang, 1995) and reflects both emotional responses as well as cognitive activity (Boucsein, 1992). GSR has been used extensively as an indicator of experience in both non-technical domains (see Boucsein, 1992 for a comprehensive review), and technical domains (e.g. Wilson and Sasse, 2000a, b; Ward and Marsden, 2003). We measured GSR using surface electrodes sewn in Velcro straps placed around two fingers on the same hand.

[1]There are a few consulting firms that specialize in observational analysis of entertainment technologies (Lazzaro, 2004); however, many researchers rely on subjective data for user preference, rather than objective observational analysis.

2.2. Cardiovascular measures

The cardiovascular system includes the organs that regulate blood flow through the body. Measures of cardiovascular activity include HR, interbeat interval (IBI), heart rate variability (HRV), blood pressure (BP), and BVP. Electrocardiograms (EKG) measure electrical activity of the heart, and HR, IBI, and HRV can be computed from EKG.

HR reflects emotional activity. It has been used to differentiate between positive and negative emotions with further differentiation using finger temperature (Winton et al., 1984; Papillo and Shapiro, 1990). To collect EKG, we placed three pre-gelled surface electrodes in the standard configuration of two electrodes on the chest and one electrode on the abdomen.

2.3. Electromyography

Electromyography (EMG) measures muscle activity by detecting surface voltages that occur when a muscle is contracted (Stern et al., 2001). In isometric conditions (no movement), EMG is closely correlated with muscle tension (Stern et al., 2001). On the face, EMG has been used to distinguish between positive and negative emotions. EMG activity over the brow region (corrugator supercilii, the frown muscle) is lower and EMG activity over the cheek (zygomaticus major, the smile muscle) and preiocular (orbicularis oculi) muscle regions are higher when emotions are mildly positive, as opposed to mildly negative (Cacioppo et al., 2000).

Smiling activity ($EMG_{smiling}$) from *zygomaticus major* activation and frowning activity ($EMG_{frowning}$) from *corrugator supercilii* activation have been able to distinguish between positive, neutral and negative valence at a rate greater than chance when viewing pictures or video as stimuli (Partala et al., 2005). We used surface electrodes on these two locations to detect voluntary smiling and frowning. The disadvantage of using surface electrodes is that the signals can be muddied by other facial muscle activity, such as talking. Needles are an alternative to surface electrodes that minimize interference, but were not appropriate for our experimental setting.

2.4. Use of physiological metrics in HCI

Physiological metrics have only recently been used in the domain of HCI. Researchers have used GSR and cardiovascular measures to examine subject response to video and audio degradations in video conferencing software (Wilson and Sasse, 2000a, b), and to investigate user response to well- and ill- designed web pages (Ward and Marsden, 2003). HRV has been used as an indicator of mental effort and stress when interacting with simulators (Vicente et al., 1987; Rowe et al., 1998) and to distinguish between attentive states of a user (Chen and Vertegaal, 2004).

Partala and Surakka (2004) and Scheirer et al. (2002) both used pre-programmed mouse delays to intentionally frustrate a computer user. Partala and Surakka measured EMG activity on the face in response to positive, negative, or no audio intervention, while Scheirer et al. applied Hidden Markov Models (HMMs) to GSR and BVP data to detect states of frustration.

Our previous work has examined physiological responses to different interactive play environments (Mandryk and Inkpen, 2004; Mandryk et al., 2006b). We showed that GSR and EMG of the jaw were higher when playing against a friend, over playing against a computer, and we found many correlations between normalized physiological activity and normalized subjective measures, including strong correlations between GSR and fun, and EMG and challenge. We also showed how physiological measures provide a rich, continuous, and objective source of information about user experience with interactive entertainment technologies. Based on these results, we believe that physiological metrics can be used to model user emotional experience when playing a game; providing continuous and objective metrics of emotion.

3. Identifying emotions

There has been a long history of researchers attempting to use physiological data to identify emotional states. William James first speculated that patterns of physiological response could be used to recognize emotion (Cacioppo et al., 2000), and although this viewpoint is too simplistic, recent evidence suggests that physiological data sources can differentiate among some emotions (Ekman et al., 1983; Levenson, 1992). For example, Picard et al. (2001) performed a feature-based recognition of eight emotional states from GSR, EMG of the jaw, BVP, and respiration over multiple days. Their algorithmic approach partially corrected for day-to-day differences, and provided an 81% accuracy on recognizing eight emotional states.

Opinions vary on whether emotions can be classified into discrete emotions (Ekman, 1999), or whether emotions exist along multiple axes (Russell et al., 1989; Lang, 1995). Both perspectives have seen limited success in using physiology to identify emotional states (Cacioppo et al., 2000). The arousal-valence space (AV space) used by Lang (1995) classifies emotions in a 2-D space defined by arousal and valence (pleasure). Using pictures as stimuli, Lang and colleagues mapped individual pictures to emotions as defined by the space.

Russell et al. (1989) also used an arousal-valence space to create the Affect Grid. Based on their circumplex model of emotion, the affect grid is a tool to quickly assess affect along dimensions in AV space. Subjects place checkmarks in the squares of the grid, as a response to different stimuli. One problem with the AV space method of classifying mood is that arousal and valence may not be independent and can impact each other. For example, Lang et al. (1993) had difficulty finding images that represent the extreme regions of the unpleasant/calm quadrant. It seems that if an image is truly unpleasant, it cannot also be calm, suggesting some interplay between these two axes (Fig. 1).

In addition to the difficulties in classifying emotions, when using physiological data sources there are methodological issues that must be addressed (Picard, 1997), and theoretical limitations to inferring significance (Cacioppo and Tassinary, 1990). Discussing these issues are beyond the scope of this paper, but a discussion can be found in (Mandryk, 2005).

4. Experimental details

We conducted a study to investigate whether we could model emotional responses to play technologies. To generate values for user emotion, we modeled the data in two parts using a fuzzy logic approach. First, we computed arousal and valence values from the normalized physiological signals of GSR, HR, $EMG_{smiling}$, and $EMG_{frowning}$. We then used these arousal and valence values to generate emotion values for boredom, challenge, excitement, frustration, and fun.

The details in this section apply to data that was collected for 12 participants. Six of the participants were used to generate the emotion models, which are described in this paper. The remaining six participants were used to validate the modeled emotions by comparing the results to reported emotions through subjective responses. The validation is discussed thoroughly in (Mandryk et al., 2006a), and presented briefly in Section 8. The experiment design is summarized in this section, while details can be found in (Mandryk et al., 2006a).

4.1. Play conditions and participants

Participants played a computer game in three conditions: against a co-located friend, against a co-located stranger, and against the computer. We were not interested in whether there was a difference between playing against a friend, a stranger, or a computer. We have observed many groups of people playing with interactive technologies, and we know that these three play conditions yield very different play experiences; rather, we were interested in whether our model of emotion could detect the differences between the conditions. Participants played NHL 2003 by EA Sports in all conditions and each play condition consisted of one 5-minute period of hockey.

Twenty-four male participants (aged 18–27) who were frequent computer users and played games frequently on either a computer or game console took part in the experiment. We wanted all of the participants to be independent subjects, statistically unrelated to any of the other participants, so we only treated one player in each pair as the participant. As such, we designed the experiment for 12 participants in 12 pairs, and we report data for 12 participants; one member of each pair. Order of the presentation of the conditions was fully counterbalanced. The stranger remained constant for all participants, and was a 29 year-old male gamer, who was instructed to match each participant's level of play to the best of his ability.

4.2. Experimental setting and protocol

The experiment was conducted in an office at Simon Fraser University. NHL 2003 was played on a Sony PS2, and viewed on a 36″ television. A camera captured both of the players, their facial expressions and their use of the controller. All audio was captured with a boundary microphone. The game output, the camera recording, and the screen containing the physiological data were synchronized into a single quadrant video display, recorded onto tape, and digitized (see Fig. 2) along with the audio from the game and the audio from the boundary microphone.

Physiological data were gathered using the ProComp Infiniti system and sensors, and BioGraph Software from Thought Technologies. Before each experimental condition, participants rested for 5 min. The resting period

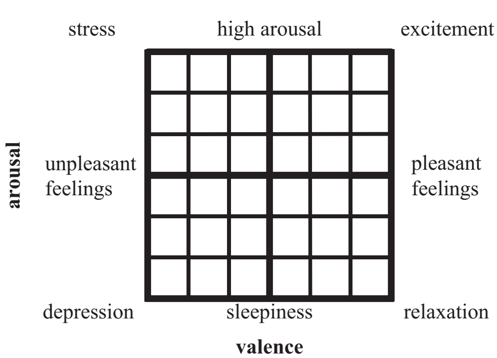

Fig. 1. Our interpretation of the Affect Grid. We changed the grid from having nine levels of arousal and valence, to having six levels of arousal and valence.

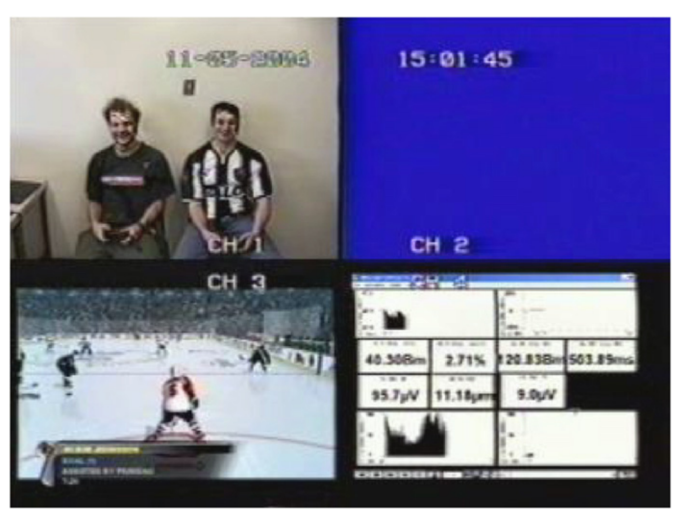

Fig. 2. Quadrant display (counter-clockwise from top left): (a) camera feed of the participants, (b) screen capture of the game, (c) screen capture of the biometrics. Audio of the game and audio of the participants' comments was included.

allowed the physiological measures to return to baseline levels prior to each play condition. Prior experiments showed that the act of filling out the questionnaires and communicating with the experimenter altered the physiological signals (Mandryk and Inkpen, 2004; Mandryk et al., 2006b). The resting periods corrected for these effects.

After each condition, the participants filled out a condition questionnaire. They were asked to consider the statement, "This condition was boring", rating their agreement on a 5-point scale with 1 corresponding to "Strongly Disagree" and 5 corresponding to "Strongly Agree". The same technique was used to rate how challenging, exciting, frustrating, and fun that particular condition was. The questionnaire was filled out online using a laptop computer to minimize the physiological effects of an interview (Mandryk and Inkpen, 2004; Mandryk et al., 2006b). After completing the experiment, subjects completed a post-experiment questionnaire asking them to decide in retrospect which condition was most enjoyable, most fun, most exciting, and most challenging.

4.3. Data analyses

The subjective data from the condition and post-experiment questionnaires were analyzed using non-parametric statistical techniques. In terms of the physiological data, EKG data were collected at 256 Hz, while GSR, respiration, and EMG were collected at 32 Hz. HR was computed at 4 Hz. Physiological data for each rest period and each condition were exported into a file. Noisy EKG data may produce heart rate (HR) data where two beats have been counted in a sampling interval or one beat has been counted in two sampling intervals. We inspected the HR data and corrected these erroneous samples. In addition, HR data were interpolated since HR was sampled at a lower frequency than the EMG or GSR signals. After interpolation, HR was smoothed using a 4 frame moving average window.

Each EMG signal was smoothed with a moving average window of 4 frames (0.125 s) (Fridlund and Cacioppo, 1986), while GSR was filtered using a 5-second window (Boucsein, 1992). We then normalized each signal into a percentage between 0 and 100. There are very large individual differences associated with physiological data, and normalizing the data is necessary to perform a group analysis. We transformed each sample into a percentage of the span for that particular signal, for each participant across all three conditions. Using GSR as an example, a global minimum and maximum GSR were obtained for each participant using all three conditions and the rest period, and the same global values were used for normalizing within each condition.

$$\text{Normalized GSR}(i) = \left(\frac{\text{GSR}(i) - \text{GSR}_{\min}}{\text{GSR}_{\max} - \text{GSR}_{\min}}\right) \times 100.$$

The same method was used to normalize the $\text{EMG}_{\text{smiling}}$, $\text{EMG}_{\text{frowning}}$, and HR data.

5. Fuzzy logic

We used normalized GSR, HR, $\text{EMG}_{\text{smiling}}$, and $\text{EMG}_{\text{frowning}}$ signals as inputs to a fuzzy logic model. To generate values for user emotion, we modeled the data in two parts. First, we computed arousal and valence values from the normalized physiological signals, then used these arousal and valence values to generate emotion values for boredom, challenge, excitement, frustration, and fun.

Fuzzy logic mimics human control logic in that it uses an imprecise but descriptive language to deal with input data, much like a human operator (Cox, 1992). Fuzzy logic systems address the imprecision of the input and output variables by defining them with fuzzy numbers and fuzzy sets that are expressed in linguistic terms (e.g., cold, warm, hot) (Tsoukalas and Uhrig, 1997). Simple, plain-language IF/THEN rules are used to describe the desired system response in terms of the linguistic variables, rather than through complex mathematical formulas. Classical sets require hard boundaries and binary memberships, whereas fuzzy sets allow for partial membership around the boundaries.

Fuzzy logic can easily represent continuous processes that are not easily broken into discrete segments, when the change of state from one linguistically-defined level to the next is not clear (Cox, 1992). In general, fuzzy logic should be used when (Cox, 1992): one or more of the control variables are continuous; when a mathematical model of the process does not exist; when high ambient noise levels must be dealt with; and when an expert can identify the rules underlying the system behavior and the fuzzy sets that represent that characteristics of each variable.

The fuzzy logic system consists of inputs, outputs, membership functions, and rules. The membership functions transform the membership of a specific element into a percentage membership in the set. It weights each input signal, defines overlap between the levels of input, and determines an output response. Membership functions can take a number of shapes; however, triangular and trapezoidal membership functions are the most common (Tsoukalas and Uhrig, 1997). The IF/THEN rules use the input membership values as weighting factors to determine their influence on the fuzzy solution sets (Cox, 1992). Once the functions are inferred, scaled, and combined, they are defuzzified[2] into a solution variable (scalar output) (Cox, 1992). Membership functions can be different for each input and output response.

There are other machine learning methods available, including neural nets. Neural nets and fuzzy systems take opposite approaches to dealing with uncertainty (Tsoukalas and Uhrig, 1997). Neural nets use precise inputs and outputs to train a generic model, while in fuzzy systems, the inputs and outputs are fuzzy and their interrelationships take the form of well-defined rules (Tsoukalas and Uhrig, 1997). One of the disadvantages of neural nets is that they

[2]We used the centroid method of defuzzification.

need substantial data that cover the entire range over which the different variables are expected to change (Tsoukalas and Uhrig, 1997). Our participants are generally happy; however, there could easily be moments when participants are bored or frustrated. We cannot guarantee that the complete span of any emotion will be covered by game playing.

Fuzzy logic systems are best used with continuous variables (Cox, 1992), like our collected physiological signals. We chose to use a fuzzy approach since there is a strong theoretical basis for the transformation from input to output; an expert can use linguistic terms to describe this transformation; we have noisy input signals; and the physiological variables are continuous.

6. Modeling arousal-valence space

The first stage was to transform the physiological signals into AV space (arousal-valence space). To generate the models, we used half of the participants (one for each play condition order), reserving the other six participants for validation of the model. To make use of the continuous nature of physiological data, we used the complete time series for each input. As such, we were able to generate a new time series of the participant's experience in AV space, rather than having only one data point for an entire condition (e.g. mean).

Our model to transform physiology to AV space had four inputs (GSR, HR, $EMG_{smiling}$, and $EMG_{frowning}$) and two outputs (arousal and valence) (see Fig. 3). Inputs were

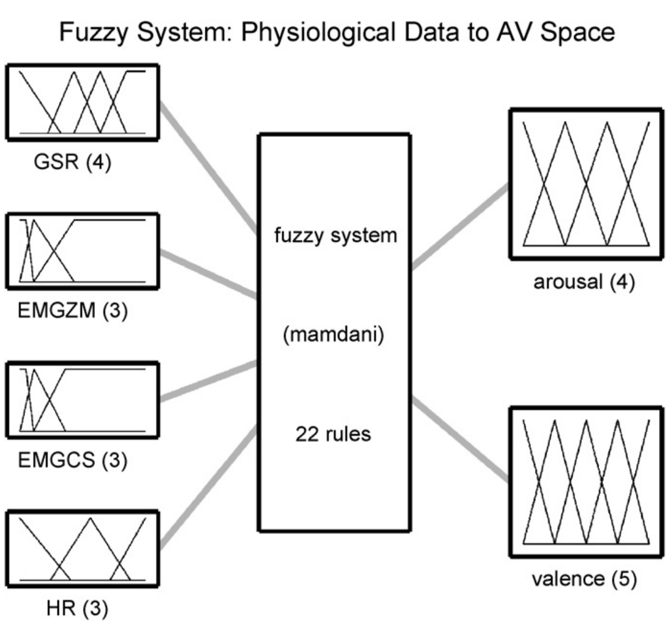

Fig. 3. Modeling arousal and valence from physiological data. The number of membership functions applied to that input or output follows the input/output labels. Within each input and output, there is a schematic representing the location and form of the membership functions. Fig. 4 through Fig. 7 show the membership functions in more detail. The system used 22 rules to transform the 4 inputs into the 2 outputs.

normalized signals (0–100), while outputs were percentages of the possible maximum (0–100) value for arousal and valence.

6.1. *Membership functions*

Membership functions were applied to the four physiological inputs and the two outputs.

6.1.1. *Input data histograms*

For each input signal, the membership functions were generated using characteristics of that particular signal over the six participants and three conditions. For each of the input signals, there are a total of 147176 samples. We generated histograms for each input, with 1000 bins, in order to have approximately 150 samples per bin. These values were chosen to maximize the number of bins while maintaining statistical relevance, and to ensure the division of value didn't exceed the precision of measurement of the samples.

Fig. 4 through Fig. 7 show how the membership functions were generated for each input signal, using the statistical characteristics of the histograms. As seen in Fig. 4, HR approaches a normal distribution. For HR, the statistical characteristics of the signal (mean, standard deviation) were used to define membership functions that suit the distribution of the input signal. The membership functions describe low, medium, and high HR activity, and were all triangular, as seen in Fig. 3.

Fig. 5 shows how GSR was distributed across the entire span, although more activity occurred in the mid and high range. As the distribution of GSR contained multiple peaks, four membership functions were used: low, mid-low, mid-high, and high. The statistical characteristics of the signal were used to determine where the membership functions were positioned. The membership functions were triangular and trapezoidal as seen in Fig. 3.

Both $EMG_{smiling}$ and $EMG_{frowning}$ were clustered towards the low end of activation, approximating lognormal distributions (see Figs. 6 and 7). For both EMG signals, three membership functions were defined, representing low, medium, and high EMG activity. Due to the statistical characteristics of a lognormal distribution, the membership functions were clustered towards the low end of activation. The medium membership function was triangular, while the low and high membership functions were trapezoidal. The trapezoids were used to remove fuzziness from the extreme values of input.

6.1.2. *Output membership functions*

Membership functions for the two outputs (arousal and valence) were distributed evenly across the entire spectrum. Arousal was defined with four memberships: low, mid-low, mid-high, and high. Valence was described by five memberships: very low, low, neutral, high, and very high. The neutral membership was introduced to accommodate the large percentage of smiling and frowning activity

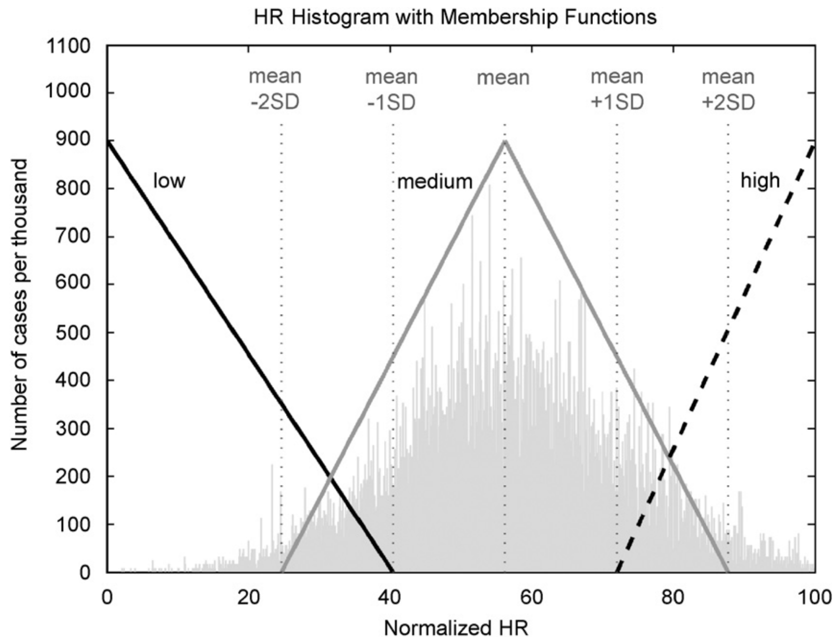

Fig. 4. Histogram of HR with statistical characteristics and three membership functions superimposed. HR approximates a normal distribution.

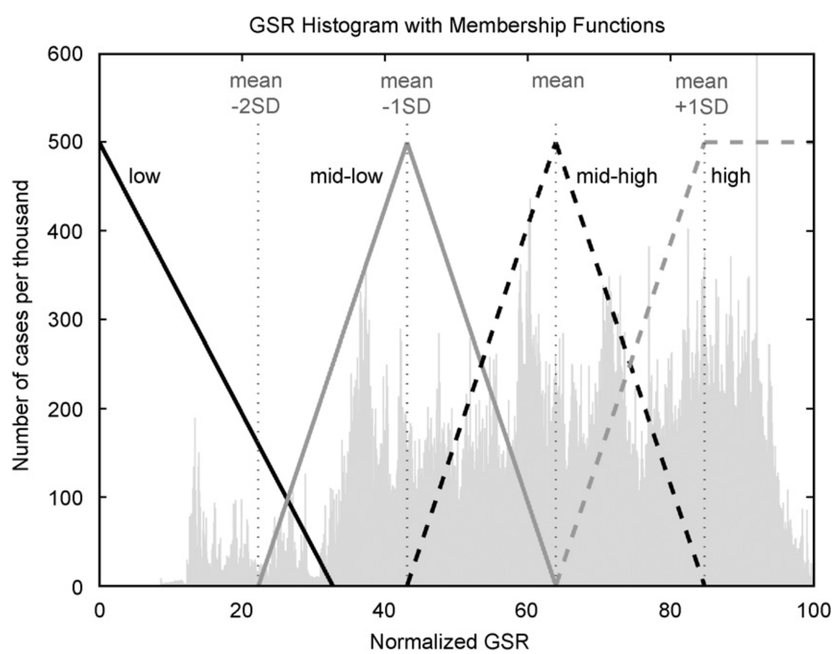

Fig. 5. Histogram of GSR with statistical characteristics and four membership functions superimposed.

that occurred at less than 5% of total activation. The output membership functions were all triangular as seen in Fig. 3.

6.2. Rules

The 22 rules were grounded in the theory of how the physiological signals relate to the psychological concepts of arousal and valence, described in Section 2. Arousal was generated from GSR and HR, while valence was generated from $EMG_{smiling}$, $EMG_{frowning}$, and HR.

GSR correlates with arousal, and increasing GSR was mapped to increasing arousal. The extreme high and low levels of GSR were modulated by HR data; if HR contradicted GSR, arousal was altered, otherwise arousal was maintained. Fig. 8 shows how GSR and HR combine through the defined rules and membership functions to generate arousal.

Since smiling activity reflects positive emotions, and frowning activity represents negative emotions, valence generally increased with increasing levels of $EMG_{smiling}$, and decreased with increasing levels of $EMG_{frowning}$. Fig. 8

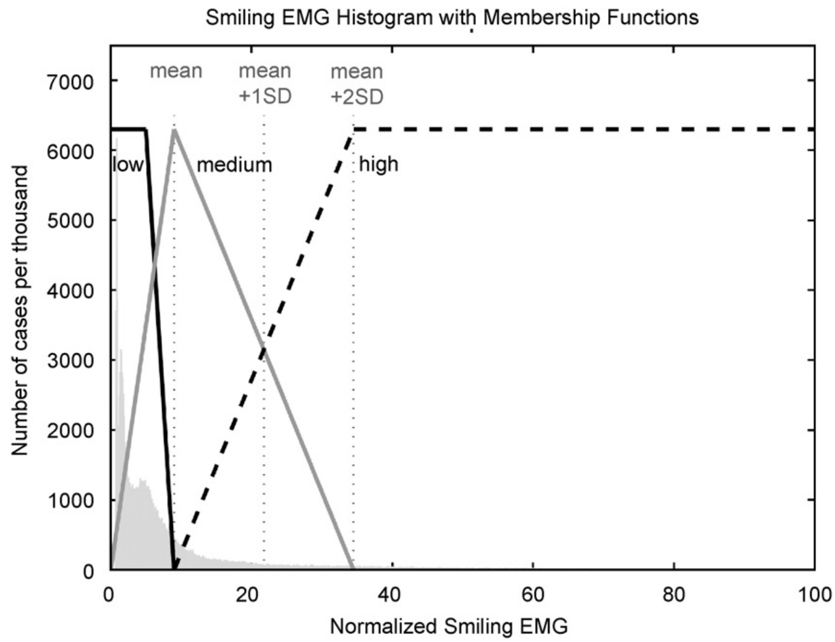

Fig. 6. Histogram of EMGsmiling with statistical characteristics and three membership functions superimposed. EMGsmiling approximates a lognormal distribution.

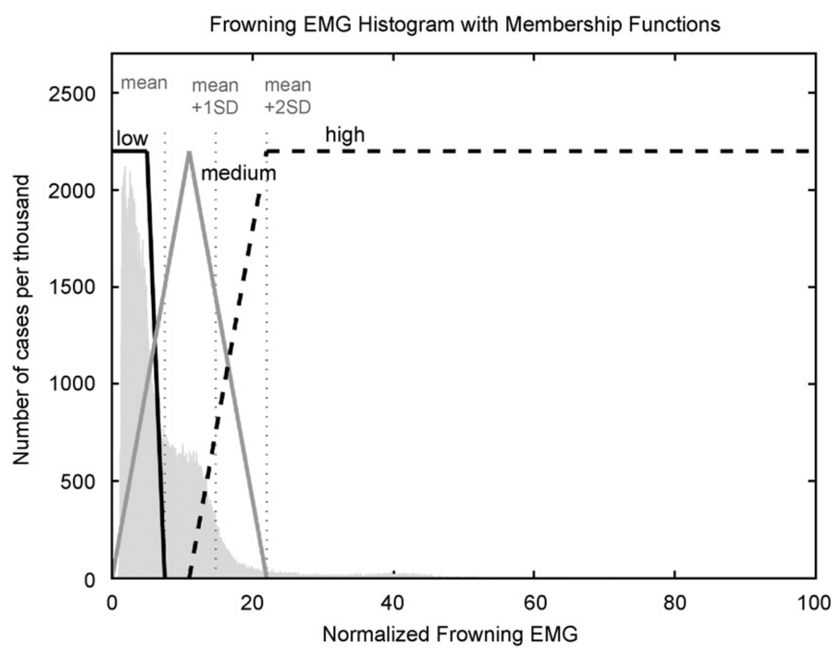

Fig. 7. Histogram of EMGfrowning with statistical characteristics and three membership functions superimposed. EMGfrowning approximates a lognormal distribution.

shows how $EMG_{smiling}$ and $EMG_{frowning}$ combine through the rules and membership functions to generate valence. Because the majority of the activation for both EMG signals occurred at less than 5%, (neutral facial expression) we would expect valence to be neutral most of the time. In addition, when $EMG_{smiling}$ and $EMG_{frowning}$ were both high, the valence output resolved to a neutral state. This type of activation would occur when participants were making a face other than smiling or frowning, and did not occur very often. When both EMG signals are low, EMG does not provide enough information to predict valence. As a result, we used HR to modulate these occurrences (see rules 18 and 19 in Appendix A). HR tends to increase with positive affect (Winton et al., 1984; Papillo and Shapiro, 1990), so when we were unable to distinguish valence for EMG alone, we used high HR values to move valence from neutral to high, and low HR values to move valence from neutral to low. The 22 rules are presented in Appendix A.

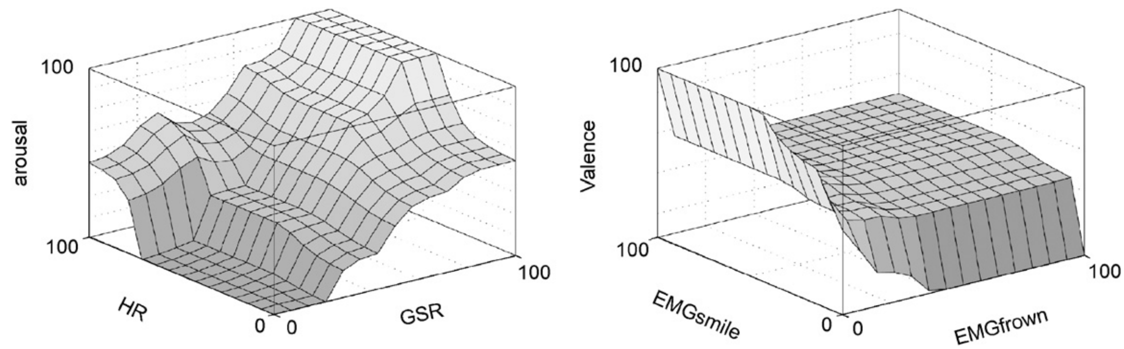

Fig. 8. Surfaces depicting how GSR, HR, EMGsmiling, and EMGfrowning are converted into arousal and valence.

Table 1
Mean arousal and valence values from the fuzzy and manual approaches

		Playing against computer		Playing against friend		Playing against stranger		Difference between conditions		
		Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	F	p	η^2
Fuzzy approach	Arousal	66.2	23.5	69.7	11.9	71.9	31.2	0.09	.919	.02
	Valence	65.5	7.4	71.9	7.1	68.1	6.2	5.70	.022	.53
Manual approach	Arousal	63.1	21.3	64.7	10.9	66.4	28.1	.97	.967	.01
	Valence	47.2	2.5	52.7	2.6	49.0	2.3	21.2	.001	.81

Play condition affected valence, but not arousal for both approaches.

6.3. Fuzzy approach results

Prior work revealed that GSR and EMG_{jaw} were higher when playing against a friend, over playing against a computer (Mandryk and Inkpen, 2004; Mandryk et al., 2006b). We would expect that arousal and valence would be higher when playing against a friend, over playing against the computer. To examine whether our model is achieving the predicted results, we looked at the mean values of arousal and valence across the play conditions.

The mean results are shown in Table 1. A repeated measures ANOVA shows that there was a significant difference in valence between the three play conditions. Post-hoc analysis revealed that valence was higher when playing against a friend than when playing against the computer ($p = .005$). There was no significant difference in arousal between the conditions, although mean arousal was greater when playing against a friend over playing against a computer.

6.4. Manual approach results

We also used a manual approach to calculate arousal and valence for each sample. The manual approach was implemented in order to confirm that the output from the fuzzy logic model was on track. For the manual calculations, we used the normalized GSR signal as the arousal metric since GSR is a linear correlate to arousal. For valence, we took normalized $EMG_{smiling}$, and subtracted $EMG_{frowning}$, and re-normalized to generate a number between 0 and 100.

The mean results are shown in Table 1. A repeated measures ANOVA shows that there was a significant difference in valence between the three play conditions. Post-hoc analysis revealed that valence was higher when playing against a friend than when playing against the computer ($p = .001$) or a stranger ($p = .005$). There was no difference in arousal between conditions.

6.5. Comparing fuzzy and manual results

We wanted to compare the arousal and valence results from the fuzzy model to the results from a manual approach using a distance metric. As such, we took the absolute difference between the fuzzy result and the manual result for each value for arousal and valence for all six participants, in all three conditions. The mean differences and maximum differences for each condition are shown in Table 2. When averaged for each condition, the mean differences between the fuzzy and manual approach were between 3% and 6% for both arousal and valence. The maximum difference between the fuzzy and manual approaches for both arousal and valence occurred in the friend condition (arousal = 20.4% and valence = 41.8%).

In all, the fuzzy approach performs in a very similar manner to the manual one. Differences were computed for every sample in the time series, (147176 samples), yet mean differences were only on the order of 5%, and maximum differences were always less than 50%. An example histogram of these differences for valence

Table 2
Mean differences between the manual approach and the fuzzy approach, separated by condition

	Playing against computer		Playing against friend		Playing against stranger		Difference between conditions		
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	F	p	η^2
Mean arousal diff. (%)	5.3	3.4	3.6	1.6	3.4	0.6	1.29	.316	.21
Mean valence diff. (%)	3.9	2.3	5.5	1.6	3.7	1.9	9.83	.004	.66
Max arousal diff. (%)	19.4	10.2	20.4	9.9	16.6	7.0	0.39	.685	.07
Max valence diff. (%)	26.6	9.6	41.8	8.4	30.3	13.4	3.27	.081	.40

Mean valence difference was higher in the friend condition than in the computer or stranger condition.

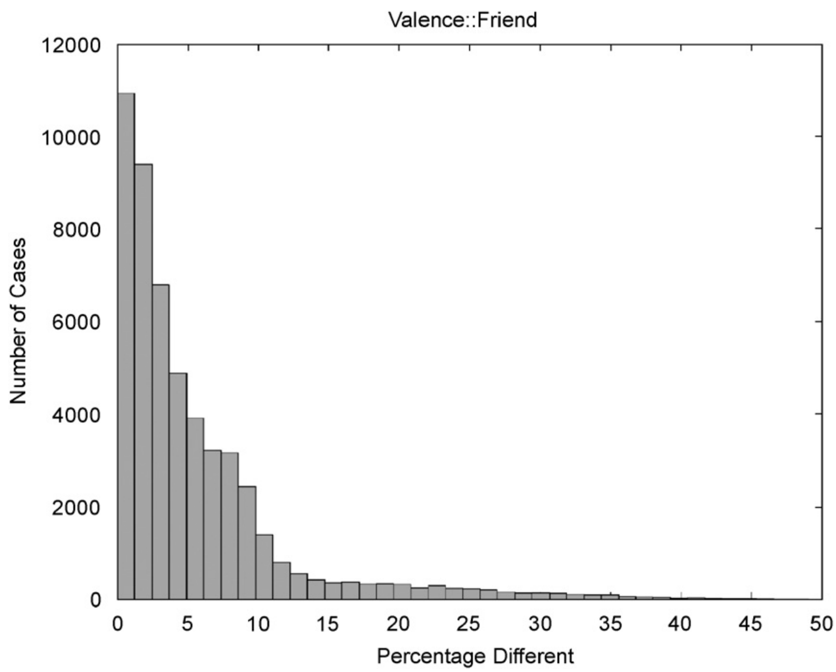

Fig. 9. A histogram reveals the total differences between the fuzzy and manual approaches for valence in the friend condition. The majority of the samples were less than 5% different.

in the friend condition is shown in Fig. 9, while other histograms of these differences can be found in (Mandryk, 2005).

We used a repeated measures ANOVA to see if the manual and fuzzy approaches were more or less comparable in each play condition. There was a significant difference in mean valence difference (see Table 2). Post hoc analysis revealed that for valence, the manual and fuzzy approaches were more similar in the stranger ($p = .010$) and computer condition ($p = .035$), than in the friend condition.

6.5.1. *AV-Space graphs*

The fuzzy and manual approaches reveal fairly similar results. In order to visualize how the two approaches differ, we generated graphs of a participant's experience in AV space over time. Traditionally, the affect grid (Russell et al., 1989) asks participants to mark an X to describe their experience in AV space. Since our approach is continuous, it is important to visualize their experience as it changed over time.

All of the participants' experiences as graphed in AV space can be found in (Mandryk, 2005). In general, we noticed that the manual approach tends to place activity in the extreme areas of AV space. Figs. 10 and 12 show Participant 16's experience in AV space when playing against a friend. The manual approach (Fig. 10) reaches the extreme positive values of both arousal and valence, whereas the fuzzy approach (Fig. 12) is less reactionary, and more moderate.

The manual approach is also more reactive to participants' facial expressions. For example, when a participant smiles, their valence increases instantly to the maximum value, whereas the fuzzy approach is a bit more moderate in evaluating valence. Figs. 11 and 13 show the AV experience for Participant 16 playing against the computer. The manual approach (Fig. 11) seems to use the neutral state as a 'home base'. Valence is generally neutral, but sometimes increases and subsequently returns to the neutral state. In contrast, the fuzzy approach (Fig. 13) is much less volatile and there is more continuity in valence throughout the experience (Figs. 10–13).

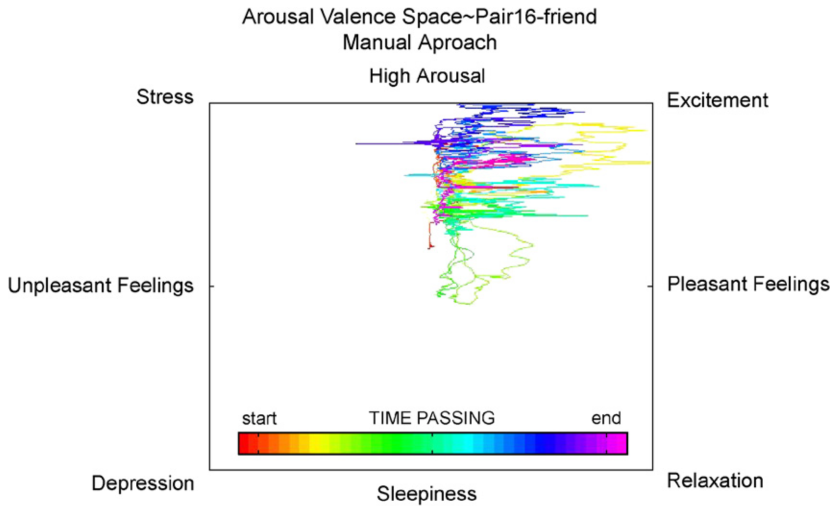

Fig. 10. The experience of Participant 16, in AV space while playing against a friend. This graph is generated using the manual approach. (The figure is reproduced in colour in the online version of the journal.)

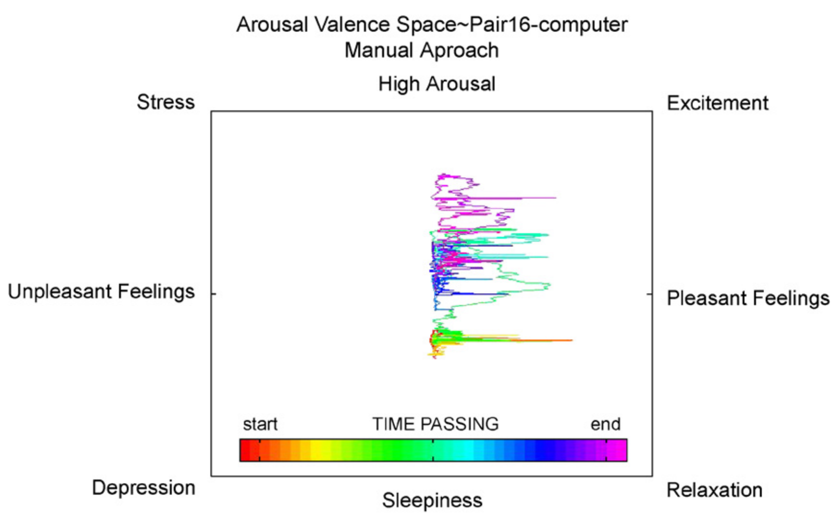

Fig. 11. The experience of Participant 16, in AV space while playing against the computer. This graph is generated using the manual approach. (The figure is reproduced in colour in the online version of the journal.)

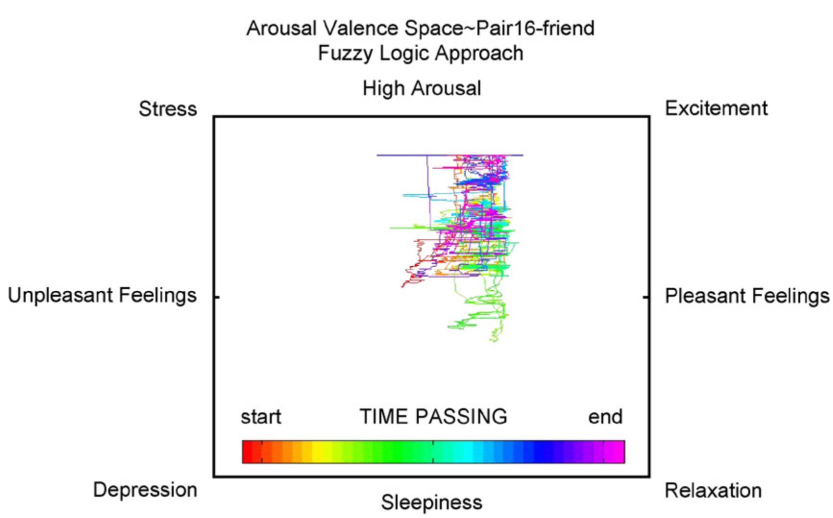

Fig. 12. The experience of Participant 16, in AV space while playing against a friend. This graph is generated using the fuzzy approach. (The figure is reproduced in colour in the online version of the journal.)

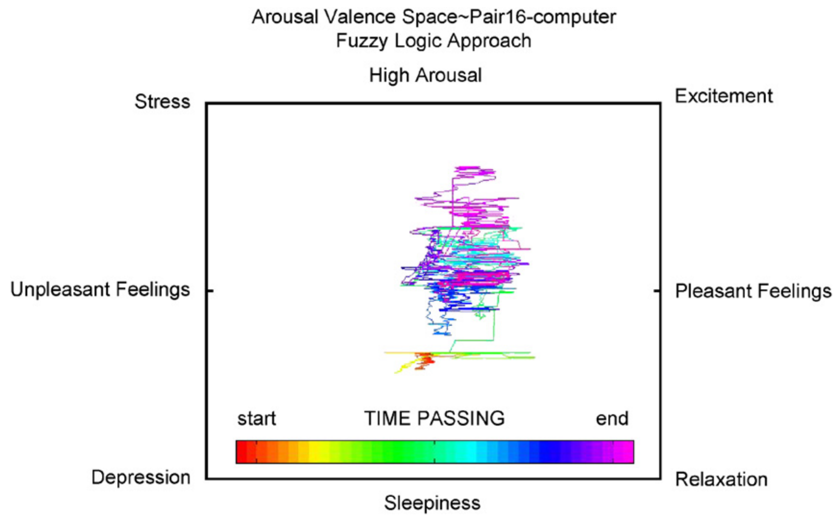

Fig. 13. The experience of Participant 16, in AV space while playing against the computer. This graph is generated using the fuzzy approach. (The figure is reproduced in colour in the online version of the journal.)

6.6. *Issues with modeling arousal and valence*

Although our AV space model is based in a theoretical understanding of the psychophysiology signals, there are some outstanding implementation issues involving the scaling of the arousal and valence axes. Our data successfully shows arousal and valence changing over time; however, the absolute positioning of this experience in AV space is difficult. In order to determine maximum arousal and valence, we used the minimum and maximum values from all three play conditions and the rest period. We determined the baseline arousal and valence values to the best of our ability, given the available data; however, the available data may not have contained accurate baseline values.

A better approach to scaling the arousal and valence axes would have been to use the IAPS (Lang et al., 1993) to baseline participants' arousal and valence. Presenting pictures from the IAPS data set, and measuring a subject's responses could provide accurate scaling information that we could use to position that subject's game-playing experience in AV space. Although informative, this process would be riddled with logistic problems since GSR is not consistent across experimental sessions (Boucsein, 1992). Baselining a participant's GSR response on one day might not apply to the following day or week. Using a variety of baselines and dynamically adjusting for the day-to-day variations (Picard et al., 2001) would be a feasible approach, requiring additional research.

In addition to scaling issues, there is also the problem that emotion-relevant ANS activity is superimposed on other physiological activity responsible for contributing to internal processes (e.g., resting and digesting, metabolic needs), and external demands (e.g., orienting, startle, and defense responses) (Levenson, 1992). As such, our results may be confounded as we attribute physiological artifacts to emotional changes. The within-subject experimental design, conducted within a single session, attempts to minimize the impact of falsely raised physiological values. Participants played in all conditions, and we compare a subject only to themselves, dealing with outstanding issues due to external factors such as ingested substances or circadian cycle.

Finally, Fig. 6 shows how $EMG_{smiling}$ was clustered towards the low end of activation, and we noticed that people did not smile much, especially when playing against the computer. Including EMG over the eye to detect involuntary smiling might enhance our model of valence.

7. Modeling emotion from AV space

The second phase of the emotion model is to use the arousal and valence information to model different emotions. To make the most of the rich, continuous physiological data, we modeled the entire AV space time series, creating continuous metrics of emotional experience. Five emotions were modeled: boredom, challenge, excitement, frustration, and fun. These are five of the seven emotions that participants rated after each play condition.

The experience states of fun and challenge are not emotions in the traditional sense; however, we felt that as they are important elements of a successful game play experience, including them in the model was a useful and beneficial research endeavor. For simplicity, they are referred to as emotions throughout the paper.

Our AV to emotion model (see Fig. 14) had two inputs (arousal and valence), and five outputs (boredom, challenge, excitement, frustration, and fun). Inputs and outputs were represented as percentages of the possible maximum.

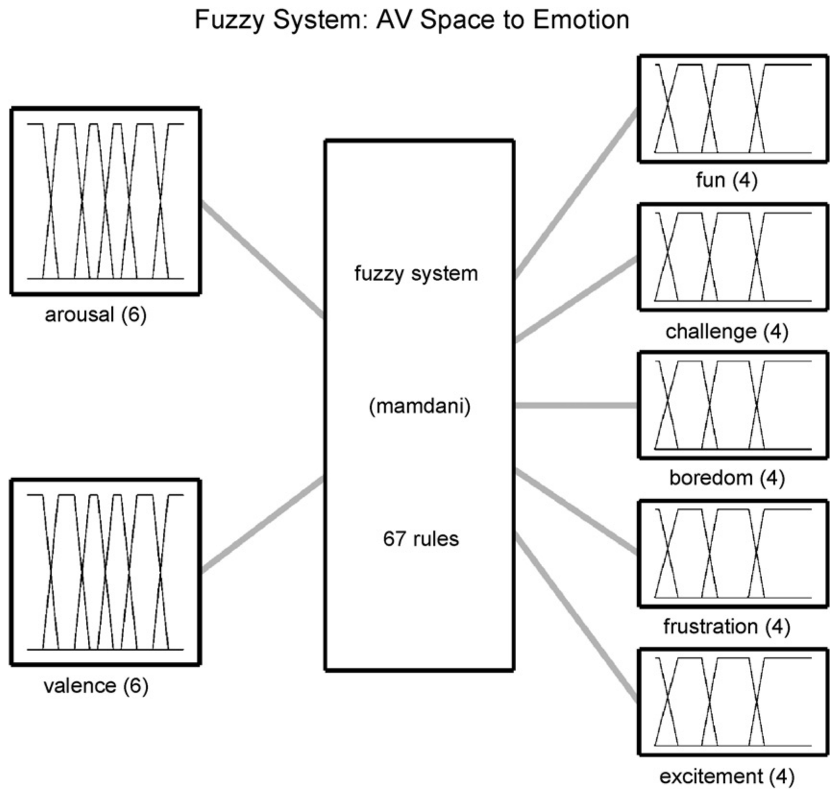

Fig. 14. Modeling emotion from arousal and valence. The number of membership functions applied to that input or output follows the input/output labels. Within each input and output, there is a schematic representing the location and form of the membership functions. All membership functions were trapezoidal, exhibited by the flat ceilings, rather than the peaked ceiling of a triangular membership function. The system used 67 rules to transform the 2 inputs into the 5 outputs.

7.1. Membership functions

The membership functions and rules for converting arousal and valence into emotion were generated using the Affect Grid, developed from the circumplex model of emotion (Russell et al., 1989). We modified the Affect Grid to have six levels of arousal and valence instead of nine levels (see Fig. 1), Using the modified Affect Grid, we mapped our arousal and valence values from the first model into a language of emotion. We represented arousal and valence in six levels: veryLow, low, midLow, midHigh, high, and veryHigh. As such, our inputs of arousal and valence used six evenly distributed membership functions. Because our mappings from arousal and valence to emotion were based on the six levels, we used trapezoidal membership functions rather than the triangular membership functions employed in the first model. The trapezoidal functions allow for a flat 'roof' on the membership function, rather than a 'point' (see Fig. 14). We wanted to remove fuzziness for the input values that were securely in the middle of any given level, and only make use of fuzziness at the boundaries between levels.

As shown in Fig. 15, we defined the five emotion outputs to have three levels: low, medium, and high, and mapped these levels onto the six levels of AV space. There are no established methods of describing levels of emotions as they vary in AV space. As such, we used guidelines from the labels on the circumplex model of emotion (Russell et al., 1989), (see Fig. 1), to define the levels of fun, challenge, boredom, frustration, and excitement (see Fig. 15). The areas in AV space where there was no mapping for a particular emotion were defined as very low for that emotion. As such, our emotion outputs were in four levels: very low, low, medium, and high (Fig. 15). As with the inputs, we used trapezoidal membership functions to only make use of fuzziness around the boundaries between levels of modeled emotion (see Fig. 14).

7.2. Rules

The rules were generated to simply map the levels of arousal and valence in Fig. 15 to the levels of fun, boredom, challenge, frustration, and fun, also shown in Fig. 15. Both arousal and valence contributed equally to the generation of boredom, challenge, excitement, frustration, and fun. The 67 rules are presented in Appendix B. Because we used data from the six subjects to iteratively generate the model, we will not present the mean results from the emotion model. See Section 8 for an analysis of the output of the emotional model for the other six subjects in the experiment.

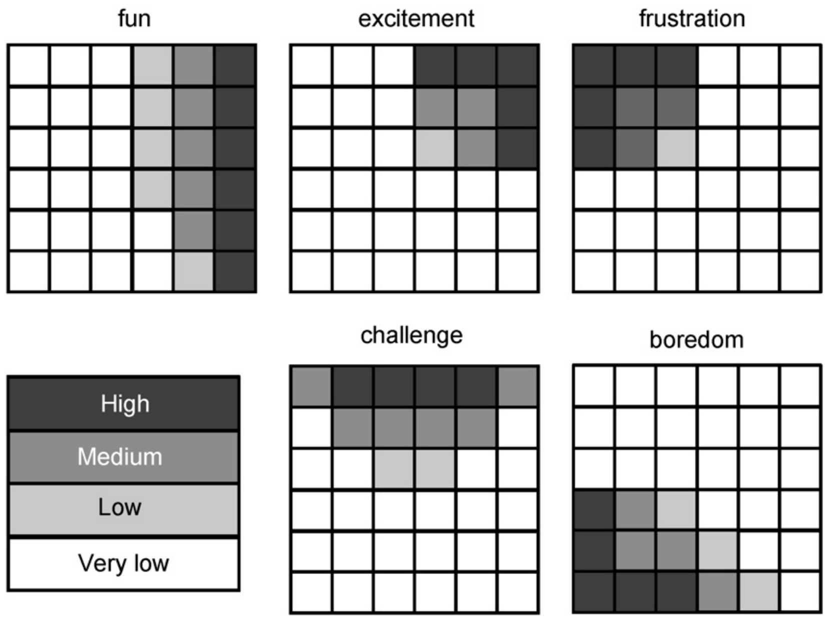

Fig. 15. Our representation of levels of emotion in arousal-valence space.

7.3. Issues with modeling emotion

The transition from AV space to the five modeled emotions was fairly straightforward. The main issue arises from the fact that there are no established guidelines for transforming levels of arousal and valence to levels of emotion in a continuous manner. We defined the membership functions and the rules to translate AV space to emotion based on the circumplex model of emotion, common sense, and our own understanding of where the five emotions exist in AV space.

In addition, there are emotions we could model which aren't easily defined in AV space, such as schadenfreude, which is taking pleasure is the misery of others. How would one use arousal and valence to describe increasing levels of pride in triumphing over adversity or gloating over the misfortune of opponents? More research needs to be conducted to determine how these emotions are described by arousal and valence before they can successfully be modeled using our fuzzy approach.

8. Using the model of emotion

To analyze the effectiveness of our model, we used data gathered from the six subjects not used in the generation of the model. Obtaining successful results using a clean set of data would show the generalizability of our model across individuals, but not across situations or applications. A complete description of the validation experiment, results, and statistics is presented in (Mandryk et al., 2006a). In addition, information on the applicability of the work to designers and other HCI professionals is discussed in (Mandryk et al., 2006a).

Data were smoothed and normalized using the previously described method (see Section 4.3). The physiological signals to AV space and AV space to emotion models were applied to the data and the time series for each emotion were averaged so that we could compare modeled emotion to the subjective responses. Although subjective responses sometimes deviate from actual experience (Marshall and Rossman, 1999; Wilson and Sasse, 2000b), we can use the reported emotions to gauge the accuracy of our model.

Mean modeled emotions (represented as a percentage) from the six new subjects were analyzed using a repeated measures MANOVA with the five emotions as dependent measures, and play condition as a within-subjects factor, while reported emotions were analyzed with a Friedman test for 3-related samples. We found that play condition impacted modeled fun and excitement (Mandryk et al., 2006a), but not boredom, challenge or frustration. Although there were no subjective differences between conditions, plotting the means reveals that there were definite trends (see Fig. 16). Furthermore, plotting the modeled emotion means reveals the same trends for boredom, excitement, and fun (see Fig. 17).

To determine how closely the modeled (objective) emotion resembled reported (subjective) emotion, we correlated the two data sources for each emotional state. The subjective and physiological emotional states were significantly correlated for fun and excitement; the same two emotional states where the model revealed significant differences across play conditions (Mandryk et al., 2006a). There was no correlation for boredom or frustration, although the same trends were present for reported and modeled boredom and frustration. The values for modeled boredom were very low and similar; the same problem

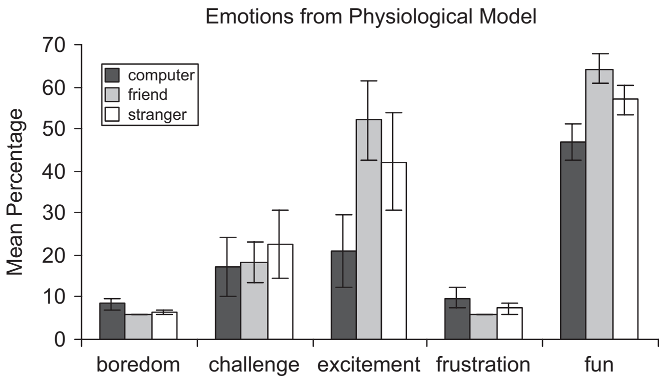

Fig. 16. Means (±SE) of the subjective reports on a 5-point scale, separated by play condition.

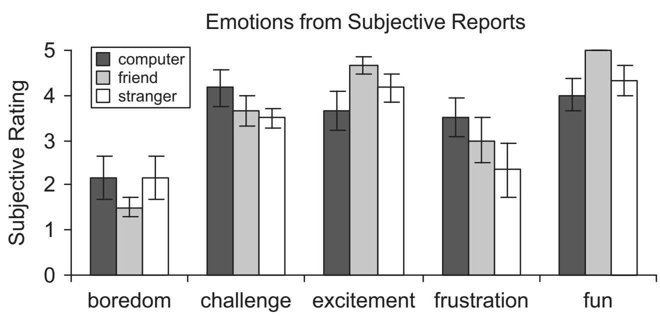

Fig. 17. Means (±SE) of modeled emotion, represented as a percentage, separated by play condition.

existed with frustration. Both of these modeled emotions suffered from issues with scaling, which are discussed in Section 8.1. There was an inverse correlation for challenge (Mandryk et al., 2006a). In modeling challenge, we assumed that a player's arousal would increase with challenge; however, upon further examination, this pattern was only true for about half of the participants, while the opposite was true for the other half. Some participants' comments revealed a strategy to attempt to relax when challenged, in order to improve their performance. Obviously, how participants handle challenge in a game is an individual strategy and additional work is required before challenge can be modeled accurately.

We also examined the subjective results from the post-experiment questionnaires. We found that maximium modeled emotion corresponded with responses for which condition was deemed the most fun in 83% of participants, most exciting for 100% of the participants, and most challenging for only 17% of the participants (Mandryk et al., 2006a). These results corroborate aforementioned mean results for each condition. Participants were not asked which condition they perceived as the most frustrating or boring.

8.1. Scaling issues

Although the modeled and reported trends between conditions are similar for most of the emotions, there are apparent differences in the relative strength of the emotions. Our model represents the emotion as a percentage of the possible maximum and minimum, given the available data. Computer games are generally fun, enjoyable experiences. Although a user may be frustrated, and may rate this frustration as fairly high on a 5-point scale, this frustration will be low when compared to the frustration experienced by getting a flat tire on the way to an important appointment. By the same logic, the boredom reported by subjects will be much lower than the boredom experienced during a really boring lecture given by a monotonous professor. We asked participants to agree with the statement "this condition was frustrating". Had we asked them to rate their response as a ratio of how frustrating it was compared to a flat tire on the way to an appointment, we probably would have seen much different subjective results. In contrast, our model takes a global approach to the scaling of emotion, so a user's frustration is given as a percentage of the maximum possible frustration, given the available data. As seen in Figs. 16 and 17, boredom, challenge, and frustration are significantly lower for modeled emotion than for reported emotion, while fun and excitement are only somewhat lower. This result is expected, since playing a computer game can be quite fun and exciting, but perhaps not as fun and exciting as riding a rollercoaster or attending a rock concert.

In addition to the scaling issues with subjective reports, Sections 6.6 and 7.3 discuss the scaling issues with the modeled emotions. Although we took a global approach to scaling, given the available data, we cannot be certain that our modeled emotions represent the percentage of the maximum value of each particular emotion exactly. We can only be certain that our values represent percentages of emotion for playing a console game. For example, had we collected GSR, HR, and facial EMG when participants were riding a rollercoaster or dealing with a flat tire, we may have seen different absolute values for our modeled emotions. Using the IAPS to scale responses in AV space, as discussed in Section 6.6, may have provided a slightly different scale.

8.2. Modeled emotion: a continuous data source

Mean modeled emotion is an objective and quantitative metric for evaluating interactive play technologies that reveals variance between conditions. In addition, modeled emotion from physiological data is very powerful as it can continuously and objectively provide a quantitative metric of user experience within a play condition. The mean values shown in Fig. 17 are derived from a time series for the five modeled emotions. As such, we can not only see the difference between conditions, but can follow the variance within a condition. Fig. 18 shows one participant's modeled frustration over time when playing against a computer, a friend and a stranger. The mean values reveal that participant three was most frustrated when playing

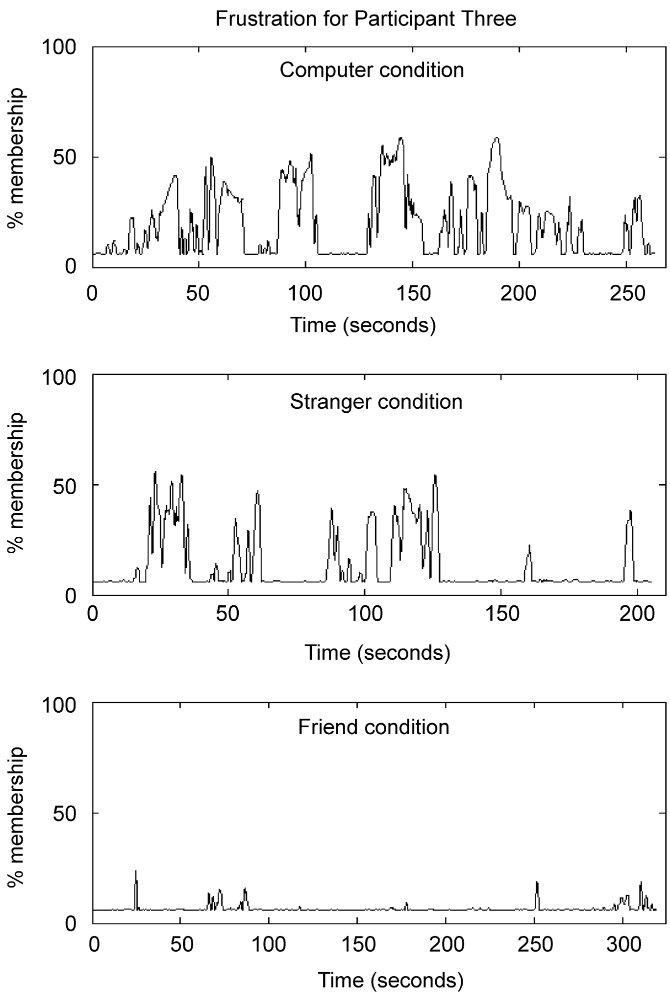

Fig. 18. Frustration for one participant in three conditions. Examining the mean output may reveal differences between conditions; however, examining the entire time series reveals how a participant's emotional state changes over time.

against the computer (mean = 19.8%), followed by playing against a stranger (mean = 13.1%), and playing against a friend (mean = 6.5%). Means alone do not tell us whether the tonic level was raised or whether there were more phasic responses. Modeled emotion pinpoints moments in time when a user's frustration was changing. This is particularly beneficial when there is no baseline or comparative condition. Researchers and developers can uncover individual moments when a user begins to get stressed, starts having fun, or becomes bored.

One of the main drawbacks to using observational analysis is the enormous time commitment associated with watching and annotating hours of video data. Continuously modeling emotion can significantly reduce the amount of time needed to perform observational analysis. By modeling emotion, researchers can look for interesting features in the emotional experience, then refer to the corresponding video to examine what events preceded the emotional reactions such as increasing boredom, increasing fun, or sustained levels of high frustration.

Researchers could also use continuous emotions to examine how the emotional experiences co-vary. Flow (Csikszentmihalyi, 1990) refers to an experience state that causes deep enjoyment, due in part to the right balance between the skill of the participant and the challenge of the activity (Csikszentmihalyi, 1990). By monitoring the change in challenge along with corresponding changes in frustration and boredom, researchers can see when players may be in danger of leaving a flow state due to an imbalance between skill and challenge. Future research could include using this information to dynamically adjust the challenge of the activity, keeping the player in a state of flow.

9. Conclusions

We used a fuzzy logic approach to transform GSR, HR, $EMG_{smiling}$, and $EMG_{frowning}$ into arousal and valence. The results from the fuzzy model were comparable to a manual approach. In addition, the results were consistent with predictions based on the results from prior experiments. A second fuzzy model was used to convert arousal and valence into five emotions: fun, challenge, boredom, frustration, and excitement. Modeled emotion was represented both as an average over a condition, and as a time series that represents an entire condition.

Mean emotion modeled from physiological data provides a method to objectively quantify user emotion when interacting with entertainment technologies. In addition, the emotion of the user can be viewed over an entire experience, revealing the variance within a condition, not just the variance between conditions. This is especially important for evaluating user experience with entertainment technology, because the success is determined by the *process* of playing, not the *outcome* of playing (Pagulayan et al., 2002). The continuous representation of emotion is a powerful evaluative tool that can be easily combined with other evaluative methods, such as video analysis. Given a time series of emotional output, researchers can identify interesting features, such as a sudden increase or decrease in an emotional state, and then investigate the corresponding time frame in a video recording. This method would drastically reduce the time required to qualitatively examine video of user interaction with entertainment technologies.

We have shown that there is great potential for using physiological metrics to model emotional experience with interactive play technologies.

Acknowledgments

Thanks to NSERC, NECTAR, EA Sports, and SFU Surrey for funding and equipment. Also, thanks to Dr. Kori Inkpen, Dr. Tom Calvert, Dr. Kelly Booth, and Dr. Lyn Bartram.

Appendix A. Rules for transforming physiological variables into arousal-valence space

The following 22 rules were used in concert with the membership functions described in Section 6 to convert GSR, HR, $EMG_{smiling}$, and $EMG_{frowning}$ into arousal and valence:

1. If (GSR is high) then (arousal is high)
2. If (GSR is mid-high) then (arousal is mid-high)
3. If (GSR is mid-low) then (arousal is mid-low)
4. If (GSR is low) then (arousal is low)
5. If (HR is low) then (arousal is low)
6. If (HR is high) then (arousal is high)
7. If (GSR is low) and (HR is high) then (arousal is mid-low)
8. If (GSR is high) and (HR is low) then (arousal is mid-high)
9. If (EMG_{frown} is high) then (valence is very low)
10. If (EMG_{frown} is mid) then (valence is low)
11. If (EMG_{smile} is mid) then (valence is high)
12. If (EMG_{smile} is high) then (valence is very high)
13. If (EMG_{smile} is low) and (EMG_{frown} is low) then (valence is neutral)
14. If (EMG_{smile} is high) and (EMG_{frown} is low) then (valence is very high)
15. If (EMG_{smile} is high) and (EMG_{frown} is mid) then (valence is high)
16. If (EMG_{smile} is low) and (EMG_{frown} is high) then (valence is very low)
17. If (EMG_{smile} is mid) and (EMG_{frown} is high) then (valence is low)
18. If (EMG_{smile} is low) and (EMG_{frown} is low) and (HR is low) then (valence is low)
19. If (EMG_{smile} is low) and (EMG_{frown} is low) and (HR is high) then (valence is high)
20. If (GSR is high) and (HR is mid) then (arousal is high)
21. If (GSR is mid-high) and (HR is mid) then (arousal is mid-high)
22. If (GSR is mid-low) and (HR is mid) then (arousal is mid-low)

Appendix B. Rules for transforming arousal-valence space into five modeled emotional states

The following 67 rules were used in concert with the membership functions described in Section 7 to convert arousal and valence into boredom, challenge, excitement, frustration, and fun:

23. If (arousal is not veryLow) and (valence is midHigh) then (fun is low)
24. If (arousal is not low) and (valence is midHigh) then (fun is low)
25. If (arousal is not veryLow) and (valence is high) then (fun is medium)
26. If (valence is veryHigh) then (fun is high)
27. If (arousal is midHigh) and (valence is midLow) then (challenge is low)
28. If (arousal is midHigh) and (valence is midHigh) then (challenge is low)
29. If (arousal is high) and (valence is midLow) then (challenge is medium)
30. If (arousal is high) and (valence is midHigh) then (challenge is medium)
31. If (arousal is veryHigh) and (valence is midLow) then (challenge is high)
32. If (arousal is veryHigh) and (valence is midHigh) then (challenge is high)
33. If (arousal is midLow) and (valence is midLow) then (boredom is low)
34. If (arousal is midLow) and (valence is low) then (boredom is medium)
35. If (arousal is low) and (valence is low) then (boredom is medium)
36. If (arousal is low) and (valence is midLow) then (boredom is medium)
37. If (arousal is midLow) and (valence is veryLow) then (boredom is high)
38. If (arousal is low) and (valence is veryLow) then (boredom is high)
39. If (arousal is veryLow) and (valence is veryLow) then (boredom is high)
40. If (arousal is veryLow) and (valence is low) then (boredom is high)
41. If (arousal is veryLow) and (valence is midLow) then (boredom is high)
42. If (arousal is midHigh) and (valence is midLow) then (frustration is low)
43. If (arousal is midHigh) and (valence is low) then (frustration is medium)
44. If (arousal is high) and (valence is low) then (frustration is medium)
45. If (arousal is high) and (valence is midLow) then (frustration is medium)
46. If (arousal is midHigh) and (valence is veryLow) then (frustration is high)
47. If (arousal is high) and (valence is veryLow) then (frustration is high)
48. If (arousal is veryHigh) and (valence is veryLow) then (frustration is high)
49. If (arousal is veryHigh) and (valence is low) then (frustration is high)
50. If (arousal is veryHigh) and (valence is midLow) then (frustration is high)
51. If (valence is veryLow) then (fun is veryLow)(challenge is veryLow)
52. If (valence is low) then (fun is veryLow)(challenge is veryLow)
53. If (valence is high) then (challenge is veryLow)(boredom is veryLow)(frustration is veryLow)
54. If (valence is veryHigh) then (challenge is veryLow) (boredom is veryLow)(frustration is veryLow)

55. If (valence is midHigh) then (boredom is veryLow) (frustration is veryLow)
56. If (arousal is veryLow) then (challenge is veryLow) (frustration is veryLow)
57. If (arousal is low) then (challenge is veryLow)(frustration is veryLow)
58. If (arousal is midLow) then (challenge is veryLow) (frustration is veryLow)
59. If (arousal is midHigh) then (boredom is veryLow)
60. If (arousal is high) then (boredom is veryLow)
61. If (arousal is veryHigh) then (boredom is veryLow)
62. If (arousal is veryLow) and (valence is midHigh) then (fun is veryLow)
63. If (arousal is low) and (valence is midHigh) then (fun is veryLow)
64. If (arousal is veryLow) and (valence is high) then (fun is low)
65. If (valence is midLow) then (fun is veryLow)
66. If (arousal is veryLow) and (valence is high) then (boredom is low)
67. If (arousal is low) and (valence is midHigh) then (boredom is low)
68. If (arousal is veryLow) and (valence is midHigh) then (boredom is medium)
69. If (arousal is veryHigh) and (valence is veryLow) then (challenge is medium)
70. If (arousal is veryHigh) and (valence is veryHigh) then (challenge is medium)
71. If (arousal is high) and (valence is low) then (challenge is low)
72. If (arousal is high) and (valence is high) then (challenge is low)
73. If (arousal is veryHigh) and (valence is low) then (challenge is high)
74. If (arousal is veryHigh) and (valence is high) then (challenge is high)
75. If (arousal is midHigh) and (valence is midHigh) then (excitement is low)
76. If (arousal is high) and (valence is midHigh) then (excitement is medium)
77. If (arousal is high) and (valence is high) then (excitement is medium)
78. If (arousal is midHigh) and (valence is high) then (excitement is medium)
79. If (arousal is veryHigh) and (valence is midHigh) then (excitement is high)
80. If (arousal is veryHigh) and (valence is high) then (excitement is high)
81. If (arousal is veryHigh) and (valence is veryHigh) then (excitement is high)
82. If (arousal is high) and (valence is veryHigh) then (excitement is high)
83. If (arousal is midHigh) and (valence is veryHigh) then (excitement is high)
84. If (arousal is midLow) then (excitement is veryLow)
85. If (arousal is low) then (excitement is veryLow)
86. If (arousal is veryLow) then (excitement is veryLow)
87. If (valence is veryLow) then (excitement is veryLow)
88. If (valence is low) then (excitement is veryLow)
89. If (valence is midLow) then (excitement is veryLow)

References

Boucsein, W., 1992. Electrodermal activity. Plenum Press, New York.

Cacioppo, J.T., Tassinary, L.G., 1990. Inferring psychological significance from physiological signals. American Psychologist 45 (1), 16–28.

Cacioppo, J.T., Berntson, G.G., Larsen, J.T., Poehlmann, K.M., Ito, T.A., 2000. The psychophysiology of emotion. In: Lewis, M., Haviland-Jones, J.M. (Eds.), Handbook of Emotions. The Guilford Press, New York, pp. 173–191.

Chen, D., Vertegaal, R., 2004. Using mental load for managing interruptions in physiologically attentive user interfaces. in the Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI 2004), pp. 1513–1516.

Cox, E., 1992. Fuzzy fundamentals. IEEE Spectrum 29 (10), 58–61.

Csikszentmihalyi, M., 1990. Flow: The Psychology of Optimal Experience. Harper Perennial, New York.

Desmet, P.M.A., 2003. Measuring emotion: development and application of an instrument to measure emotional responses to products. In: Blythe, M., Overbeeke, K., Monk, A.F., Wright, P. (Eds.), Funology: From Usability to Enjoyment. Kluwer Academic Publishers, Dordrecht, pp. 111–123.

Desurvire, H., Caplan, M., Toth, J.A., 2004. Using heuristics to evaluate the playability of games. In: the Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI 2004), pp. 1509–1512.

Ekman, P., 1999. Basic emotions. In: Dalgleish, T., Power, M. (Eds.), Handbook of Cognition and Emotion. John Wiley & Sons Ltd., Sussex.

Ekman, P., Levenson, R.W., Friesen, W.V., 1983. Autonomic nervous system activity distinguishes among emotions. Science 221 (4616), 1208–1210.

Fisher, C., Sanderson, P., 1996. Exploratory data analysis: exploring continuous observational data. Interactions 3, 25–34.

Fridlund, A.J., Cacioppo, J.T., 1986. Guidelines for human electromyographic research. Psychophysiology 23, 567–589.

Fulton, B., Medlock, M., 2003. Beyond focus groups: Getting more useful feedback from consumers. Paper presented at the Game Developer's Conference, San Jose, CA, March 6–8, 2003.

Lang, P.J., 1995. The emotion probe. American Psychologist 50 (5), 372–385.

Lang, P.J., Greenwald, M.K., Bradley, M.M., Hamm, A.O., 1993. Looking at pictures: affective, facial, visceral, and behavioural reactions. Psychophysiology 30, 261–273.

Lazzaro, N., 2004. Why we play games: 4 keys to more emotion. Paper presented at the Game Developer's Conference, San Jose, CA, March 22–26, 2004.

Levenson, R.W., 1992. Autonomic nervous system differences among emotions. American Psychological Society 3 (1), 23–27.

Mandryk, R.L., 2005. Modeling user emotion in interactive play environments: A fuzzy physiological approach. Ph.D. Thesis, Simon Fraser University, Burnaby.

Mandryk, R.L., Inkpen, K., 2004. Physiological indicators for the evaluation of co-located collaborative play. In: Proceedings of Computer Supported Cooperative Work (CSCW 2004), pp. 102–111.

Mandryk, R.L., Atkins, M.S., Inkpen, K.M., 2006a. A continuous and objective evaluation of emotional experience with interactive play environments. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2006), pp. 1027–1036.

Mandryk, R.L., Inkpen, K., Calvert, T.W., 2006b. Using psychophysiological techniques to measure user experience with entertainment technologies. Behaviour and Information Technology (Special Issue on User Experience) 25 (2), 141–158.

Marshall, C., Rossman, G.B., 1999. Designing Qualitative Research. 3rd ed. Sage, Thousand Oaks.

Norman, D.A., 2002. Emotion and design: attractive things work better. Interactions 9, 36–42.

Pagulayan, R.J., Keeker, K., Wixon, D., Romero, R., Fuller, T., 2002. User-centered design in games. In: Jacko, J., Sears, A. (Eds.), Handbook for Human–Computer Interaction in Interactive Systems. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, pp. 883–906.

Papillo, J.F., Shapiro, D., 1990. The cardiovascular system. In: Cacioppo, J.T., Tassinary, L.G. (Eds.), Principles of Psychophysiology: Physical, Social, and Inferential Elements. Cambridge University Press, Cambridge, pp. 456–512.

Partala, T., Surakka, V., 2004. The effects of affective interventions in human–computer interaction. Interacting with Computers 16, 295–309.

Partala, T., Surakka, V., Vanhala, T., 2005. Person-independent estimation of emotional experiences from facial expressions. In: The Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI 2005), pp. 246–248.

Picard, R., 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.

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

Rowe, D.W., Sibert, J., Irwin, D., 1998. Heart rate variability: Indicator of user state as an aid to human–computer interaction. In: The Proceedings of the Conference on Human Factors in Computing Systems (CHI 1998), pp. 480–487.

Russell, J.A., Weiss, A., Mendelsohn, G.A., 1989. Affect grid: a single-item scale of pleasure and arousal. Journal of Personality and Social Psychology 57 (3), 493–502.

Scheirer, J., Fernandez, R., Klein, J., Picard, R., 2002. Frustrating the user on purpose: a step toward building an affective computer. Interacting with Computers 14 (2), 93–118.

Stern, R.M., Ray, W.J., Quigley, K.S., 2001. Psychophysiological Recording. 2nd ed. Oxford University Press, New York.

Sweetsner, P., Wyeth, P., 2005. Gameflow: a model for evaluating player enjoyment in games. ACM Computers in Entertainment 3 (3) Article 3A.

Tsoukalas, L.H., Uhrig, R.E., 1997. Fuzzy and Neural Approaches in Engineering. John Wiley & Sons, Inc., New York.

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.

Ward, R.D., Marsden, P.H., 2003. Physiological responses to different web page designs. International Journal of Human-Computer Studies 59 (1/2), 199–212.

Wiberg, C., 2003. A measure of fun: extending the scope of web usability. Ph.D. Thesis, Umeå University, Umeå, Sweden.

Wilson, G.M., Sasse, M.A., 2000a. Do users always know what's good for them? Utilizing physiological responses to assess media quality. In: The Proceedings of HCI 2000: People and Computers XIV—Usability or Else! (HCI 2000), pp. 327–339.

Wilson, G.M., Sasse, M.A., 2000b. Investigating the impact of audio degradations on users: Subjective vs. Objective assessment methods. In: The Proceedings of OZCHI 2000: Interfacing Reality in the New Millennium (OZCHI 2000), pp. 135–142.

Winton, W., Putnam, L., Krauss, R., 1984. Facial and autonomic manifestations of the dimensional structure of emotion. Journal of Experimental Social Psychology 20, 195–216.