Boredom, Engagement and Anxiety as Indicators for Adaptation to Difficulty in Games

Guillaume Chanel[1], Cyril Rebetez[2], Mireille Bétrancourt[2], Thierry Pun[1]

Computer Science Departement – University of Geneva
Route de Drize 7
CH - 1227 Carouge, Switzerland
Phone: (+41) 22 379 01 83
guillaume.chanel@cui.unige.ch

TECFA – University of Geneva
Route des Acacias 54
CH - 1227 Carouge, Switzerland
Phone: (+41) 22 379 93 72
cyril.rebetez@tecfa.unige.ch

ABSTRACT
This paper proposes an approach based on emotion recognition to maintain engagement of players in a game by modulating the game difficulty. Physiological and questionnaire data were gathered from 20 players during and after playing a Tetris game at different difficulty levels. Both physiological and self-report analyses lead to the conclusion that playing at different levels gave rise to different emotional states and that playing at the same level of difficulty several times elicits boredom. Emotion assessment from physiological signals was performed using a SVM (Support Vector Machine). An accuracy of 53.33% was obtained on the discrimination of three emotional classes, namely boredom, anxiety, engagement.

Categories and Subject Descriptors
J.4 [**Computer applications**]: Social and behavioral sciences - *Psychology*

General Terms
Measurement, Experimentation, Human Factors.

Keywords
Emotion assessment, gaming engagement, physiological signals processing and classification, SVM, GSR.

1. INTRODUCTION
Due to their capability to present information in an interactive and playful way, computer games have gathered increasing interest as tools for education and training [14]. Games are also interesting from a human-computer interaction (HCI) point of view, because they are an ideal ground for the design of new ways to communicate with the machine. Affective computing [13] has opened the path to new types of human-computer interfaces that adapt to affective cues from the user. As one of the main goals of games is to provide emotional experiences such as fun and excitement, affective computing is a promising area of research to enhance game experiences.

Two main objectives for emotion assessment in games are: (*i*) evaluating the game from a user-centered perspective [11], (*ii*) maintaining involvement of the player by adapting game difficulty or content to induce particular emotional states [4, 16]. The first objective can be accomplished by using subjective and objective evaluations like self-reports and observational analysis. However, Mandryk *et al* [11] proposed to use online automatic assessment of emotions to avoid biased self-reports and extensive analysis of observational data. For the second objective, automatic assessment of emotions is mandatory for the game to adapt in real time to the feelings and involvement of the player, without interrupting his or her gaming experience. The present work focuses on the later objective although it can also have implications on the first.

Since emotion can be expressed via several channels various modalities can be used to analyze and determine emotional states of a player. In this study physiological signals from both the peripheral and central nervous system were used for this purpose. The choice of using central signaling in addition to peripheral data was done because the implication of the brain in emotional processes is now fully demonstrated and a number of studies have shown the usability of peripheral signals for emotion recognition,. As to the choice itself of physiological signals, we believe that they cannot be easily faked as is the case with speech or facial expressions. Since this is a work in progress only results from the peripheral signals are presented.

Games can elicit a lot of different emotional states but not all of them are useful to maintain involvement in the game. Many representations of the player affective state have been used in previous studies like anxiety, frustration, engagement, distress and the valence / arousal space [7, 16]. According to emotion and flow theories [5, 6] strong involvement in a task occurs when the skills of an individual meets the challenge of a task (Figure 1). Too much challenge would raise anxiety but not enough would induce boredom. In a game, this change of state can occur due to two main reasons. First, the difficulty is increased because of the

MindTrek'08, October 7–9, 2008, Tampere, Finland.

progression in the different levels but too fast compared to the competence increase of the player (potentially giving rise to anxiety) or the competence of the player have increased while the game remained at the same difficulty (potentially giving rise to boredom). In both cases, the challenge should be corrected to maintain a state of pleasure and involvement, showing the importance of having games that increase their difficulty according to the competence and emotions of the player. Based on this theory, we defined three emotional states of interest that corresponds to three well separated areas of the valence arousal space: boredom (negative-calm), engagement (positive-excited) and anxiety (negative-excited).

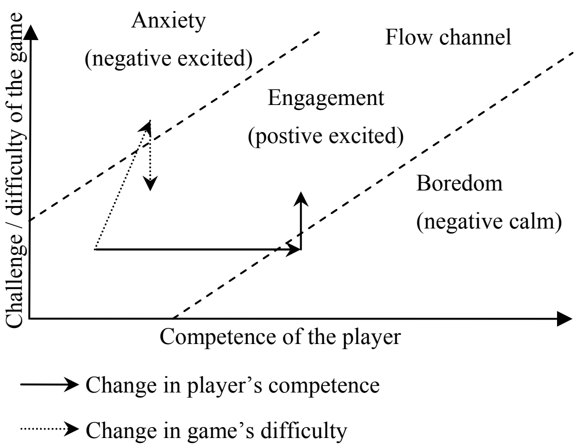

Figure 1. Flow chart and the suggested automatic adaptation to emotional reactions

This work attempts to verify the validity and usefulness of the three definied emotional sates by using a Tetris game where the challenge is modulated by changing the level of difficulty. Self-reports as well as physiological activity were obtained from players. Using those data, three hypotheses were tested:

- H1: playing at different levels of difficulty will give rise to different emotional states;
- H2: those emotional states (and the underlying conditions) can be assessed using central and peripheral signaling;
- H3: as the skill increases, the player will switch from the engagement state to the boredom state.

The protocol used to gather self-reported and physiological data as well as methods used to analyse those data are described in section 2. Results of those analysis and validation of the different hypothesis are presented in section 3. Finally, this study open the way to new reasearch paths in section 4.

2. METHODS

2.1 Data collection

In order to test hypotheses, a gaming protocol was designed for acquiring physiological signals and gathering self-reported data. The Tetris game was chosen in this experiment for the following reasons: it is easy to control the difficulty of the game (speed of falling blocks), it is a widely known game so that we could expect to gather data from players with different skill levels (which occured), and it is playable using only one hand. The last reason is mandatory since the other hand is used for some of the data acquisition sensors.

20 participants (mean age: 27, 13 males, all right handed) took part in this study. After signing a consent form, each participant played Tetris several times to determine the game level where he/she reported engagement. This was done by using the threshold method, starting from a low level and progressively increasing it until engagement was reported or starting from a high level and decreasing it. The average of the obtained levels was then considered as the participant skill level. Depending on this skill level, three experimental conditions were determined: medium condition (game difficulty equal to the player's skill level), easy condition (lower difficulty, computed by subtracting 8 levels of difficulty from the player's skill level), and hard condition (higher difficulty, computed by adding 8 levels).

Participants were then equipped with several sensors to measure their peripheral physiological activity: a GSR (Galvanic Skin Response)sensor to measure skin resistance, a plethysmograph device to record relative blood pressure, a respiration belt to estimate abdomen extension and a temperature sensor to measure palmar changes in temperature. Those sensors are known to measure signals that are related to particular emotional activation as well as usable for emotion detection [2, 3, 8]. In addition, an EEG system was used to record central signaling from 14 participants. As demonstrated in other studies, EEG's can help to assess emotional states and is also useful to provide an index of task engagement. However, their analysis is still work in progress; only results from peripheral signaling are presented in this paper. Our preliminary results related to the fusion of peripheral signals with EEG's for classifying emotional states however show an improvement with respect to using peripheral signals only [3]. All signals were recorded at a 1024Hz sampling rate using the Biosemi Active 2 acquisition system.

Once equipped with the sensors, participants played 6 sessions of Tetris. Each session lasted for 5 minutes and corresponded to one of the three experimental conditions. Each experimental condition was applied twice and in a random order to account for side effects of time in questionnaires and physiological data. The goal of participants was to perform the highest score as possible. To motivate them toward this goal, a prize was offered to participants with the highest score in three competence categories. After each session, participants filled in a questionnaire with 30 questions related to both the emotions they felt and their level of involvement in the game. The answer to each question was given on a 7 points Likert scale. Additionally, participants rated their emotion in the valence-arousal space using self-assessment manikin [12] scales.

2.2 Data analysis

To test for hypothesis H1, a factor analysis was performed on the questionnaires to find the axes of maximum variance, and the answers were projected in this new space. An ANOVA test was applied to those new variables to check for differences in distributions of judgment for the different conditions.

Prior to analyze physiological data it is necessary to preprocess signals and extract features that are known to be related to emotion activation. All signals were first filtered by a moving average filter to remove noise. For this purpose we used filters of length 512 for GSR, 128 for blood pressure, and 256 for

Table 1. Description of features extracted from peripheral signals and results of the anova test for the 3 difficulty levels. * stands for a p-value < 0.5 and ** for a p-values < 0.1 The last column indicate the trend of the median value for the 3 difficulty levels. For instance ↘ indicate that the median deacrease from the easy to the high level while →↗ indicate its stability between the easy and medium condition and an increase to the hard condition

Peripheral signal	Extracted features	Comments	F-value	p-value	Trend of median
GSR	Mean skin resistance	Estimate of general arousal level	5.5	< 0.01 **	↘
	Mean of derivative	Average GSR variation	2.8	< 0.1 *	↘↗
	Mean of derivative for negative values only	Average decrease rate during decay time	3.1	< 0.05 **	↘
	Proportion of negative samples in the derivative vs. all samples	Importance and duration of the resistance fall	6.7	< 0.01 *	↗
Blood pressure	Mean value	Estimate of general pressure	1.2	0.31	
	Standard deviation	Estimate of blood pressure (in)stability	0.7	0.5	
Heart rate	Mean of heart rate	-	3.7	< 0.05 **	↗
	Mean of heart rate derivative	Estimations of heart rate variability	2.9	< 0.1 *	→↗
	Standard deviation of heart rate	Average heart rate variation	0.6	0.55	
Respiration	Main frequency computed as the frequency having the highest energy.	-	0.5	0.6	
	Standard deviation	Variation of the respiration signal	5	< 0.01 **	→↗
	Maximum value minus minimum value	Dynamic range	1	0.35	
Temperature	Mean value	Estimate of general temperature	6.7	< 0.01 **	↘
	Average derivative	Average temperature variation	10	< 0.01 **	↘

respiration. Those different lengths were chosen to remove high frequencies without corrupting oscillations of interest. A baseline was also removed in all signals to account for changes of physiological variables due to time and movements during the feeling of questionnaires. The baseline was computed for each session, for GSR and blood pressure signals only. It was defined as the average of signals taken between 1 min 30 s and 30 s before the beginning of the session.

Table 1 presents the list of features that were computed from the different signals on the 5 min duration of the session. GSR signals have two different components: a tonic level and a phasic responses. The phasic response can occurs when there is a spontaneous event of high arousal. The extracted features try to characterize both components. It was also shown that the average GSR value is correlated with arousal [8, 10]. The plethysmograph is a device that is clipped on a finger to continuously measure relative blood pressure in the vessels. Since at each heart beat there is an increase of blood pressure, heart rate was computed by finding the local maxima of the signal. Blood pressure and heart rate variability are variables that have significant correlation with defensive reactions [8] and pleasantness of stimuli [10]. The increase of blood pressure during fear and anger is one of the most consistent findings in emotion research from autonomic activity [17]. In [15] an increase in heart rate for many basic emotions was observed. The respiration signal is obtained by a belt that measures the expansion of the abdomen related to the quantity of inspired and expired air. Slow respiration is linked to relaxation while irregular rhythm, quick variations, and cessation of respiration correspond to more aroused emotions like anger or fear [9, 15]. Laughing is also known to affect the respiration pattern by introducing high-frequency fluctuations of the recorded signal. To capture those fluctuations several features are therefore used (Table 1). To obtain the main respiration frequency it is necessary to compute the power spectral density. This was done by using the Welch algorithm with a time window of 20 s. Palmar skin temperature was also recorded since it changes in different emotional states [10]. Extracted features were chosen to represent the variations and tonic level of temperature.

As for the questionnaire, the physiological features were subject to an ANOVA test to search for differences in activations in the different conditions and check the relevance of each feature for emotion assessment. Hypothesis H2 was analyzed by training an RBF SVM (Radial Basis Function Support Vector Machine) [1] classifier on relevant physiological features to recover the three original conditions. For each participant the classifier was trained using features of the other participants; accuracy was then computed by applying the trained model on the physiological data of the tested participant. The gamma parameter of the RBF SVM was chosen to maximize accuracy on the training set. Since the classifier is tested on the data of participants that are not present in the training set, this method allows evaluating the performance of the classifier in the worst case where the model is not user-specific, i.e. no information about the specificity of the user's physiology is required for emotion assessment, except a baseline recording of 1 min.

Hypothesis H3 was tested by focusing on the data of the two sessions corresponding to the medium condition where the participant is expected to be engaged. Both physiological and questionnaire data were analyzed using a pairwise t-test to verify if there was a decrease of engagement from the first to the second session.

3. RESULTS AND DISCUSSION

The factor analysis applied on the questionnaires allowed for the definition of two components accounting for 55.6% of the variance. Based on the weights, the factors were interpreted as positive valence and arousal. The first component was positively correlated with questions related to pleasure, interest, motivation and focus while the second was positively correlated with question corresponding to levels of excitation and pressure and negatively correlated with calm and control levels. The ANOVA test, applied on the data projected in this two component space (see Figure 2), showed that participants felt lower valence for the easy and hard conditions than for the medium one ($F=46$, $p<0.01$). Differences in the three arousal distributions demonstrated that increasing difficulty led to higher reported arousal ($F=232$, $p<0.01$). This demonstrates that an adequate level of difficulty is necessary to engage players in the game and that the different playing difficulties successfully elicited different emotional states with different levels of valence and arousal.

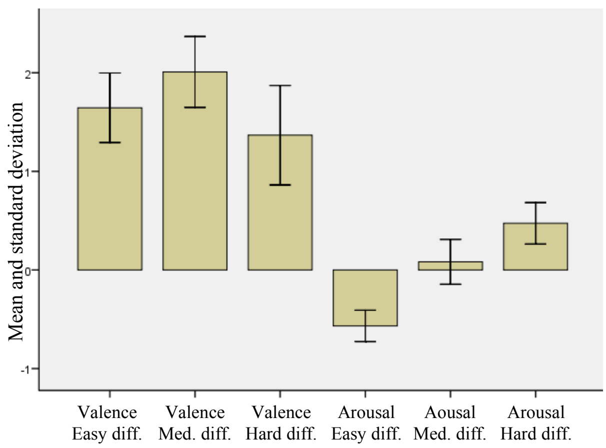

Figure 2. Mean and standard deviation of judgements for each axis of the two component space (valence and arousal) and the different difficulties (diff.): easy, medium (med.) and hard.

Results of the ANOVA test for the peripheral features are presented in Table 1. A decrease in GSR and temperature average values was observed when increasing game difficulty. An increase of heart rate was also found. Those results are known to be related to an increase of arousal ([8, 10]) and thus confirm the results obtained from the analysis of questionnaires. As can be seen from Table 1 a total of nine features where found to have significantly different distributions among the three difficulties. One feature of interest is the average GSR derivative because values of this feature are lower for the medium condition than for the two others, showing that this condition can elicit particular peripheral activation.

By using the nine relevant peripheral features to train and test the RBF SVM, an accuracy of 53.33% was achieved for discriminating between 3 classes: boredom, engagement, anxiety. Table 2 presents the confusion matrix for the 3 classes: it can be seen that the boredom condition, followed by the anxiety condition, were correctly classified while samples from the engagement condition tends to be classified as anxious or bored samples. This is not surprising since this condition lies in between the others. Notice that 30% of the samples belonging to the anxiety class are classified as bored samples, this can be due to fact that some participants completely disengaged from the task because of its difficulty reaching am emotional state close to boredom.

A pairwise t-test on the variables of the questionnaire for the medium condition showed a negative trend in the questions "I had pleasure to play" ($t=-1.8$, $p<0.09$) and "I had to adapt to the interface" ($t=-3$, $p<0.06$). From peripheral signals, an increase in mean GSR ($t=3$, $p<0.01$) and average derivative of temperature ($t=2.3$, $p<0.04$) was found as well as a decrease in average heart rate ($t=-1.9$, $p<0.08$). Those results are indicative of a decrease of arousal and pleasure while playing twice in the same condition, thus supporting the third hypothesis. The result obtained for the question "I had pleasure to play" gives a cue that this decrease could be due to an increase of player's competence. However competence was not measured with reliable indicators to confirm this possibility. In any case, those results demonstrate the importance of having automatic adaptation of game's difficulty even if the challenge of the game remains the same.

Table 2. Confusion matrix for the three classes.

Classified / True	Easy (Boredom)	Medium (Engagement)	Hard (Anxiety)
Easy (Bored.)	72.5%	20.0%	7.5%
Medium (Eng.)	37.5%	20.0%	42.5%
Hard (Anxiety)	29.0%	2.6%	68.4%

4. CONCLUSION

This study investigated the possibility and the interest of emotion assessment from physiological signals to adapt the difficulty of a game. The results obtained from self-reports and physiological analysis showed that playing a Tetris game at different levels of difficulty, chosen according to the skill of the player, gave rise to different emotional states that can be defined as boredom, engagement and anxiety. At least two of those emotional states could be detected from physiological signals with good accuracy. It was shown that the engagement of a player can decrease if the game difficulty does not change. This demonstrates the interest of modulating the difficulty of the game according to the emotions of the player.

Future work will focus on the improvement of the detection accuracy. Reducing participant variances as well as session variances by improving the current baseline is certainly a step in that direction. The analysis of EEG signals should also be continued. Preliminary results indicate that fusion of the EEG and peripheral signals will certainly increase the classification accuracy. Another question of interest is to determine the number of classes to be detected. Since boredom and anxiety are detected with high confidence it might be enough to use those two classes for adaptation to the game difficulty. Moreover, from the observation of Figure 1, one can conclude that it is more interesting to adapt the difficulty of the game solely based on the increase of competence because it leads to a stronger change of state in the flow chart and stimulates learning. In this case only the detection of boredom is of importance to modulate difficulty. This also implies to more clearly define what are the relations between emotions, competence and learning wich is one of our research interest. Another issue is the problem of disengagement in hard conditions. In this case, the adaptive game we propose would increase the level of difficulty since the detected emotion would be boredom, which is not the proper decision to take. A solution could be to use contextual information such as the current level of difficulty and the direction of the last change in difficulty (i.e. increase or deacrease). Analysis of the physiological signals should also be conducted on the basis of the events in the game and not only for the complete 5 min of a session since emotions (e.g. anxiety) can occur in any of the sessions. Finally, the last step will be to implement an adaptive Tetris game and verify that it is more fun and enjoyable than the standard one.

5. REFERENCES

[1] Bishop, C.M. Pattern recognition and machine learning. Springer, 2006.

[2] Chanel, G., Ansari-Asl, K. and Pun, T., 2007. Valence-arousal evaluation using physiological signals in an emotion recall paradigm. in IEEE SMC and International Conference on Systems, Man and Cybernetics, Smart cooperative systems and cybernetics: advancing knowledge and security for humanity, (Montreal, Canada, 2007).

[3] Chanel, G., Kronegg, J., Grandjean, D. and Pun, T. Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals in Multimedia Content Representation, Classification and Security B. Gunsel, A.K.J., A. M. Tekalp, B. Sankur ed., Springer LNCS, Istanbul, Turkey, 2006, 530-537.

[4] Chen, J. 2007. Flow in games (and everything else) - A well-designed game transports its players to their personal Flow Zones, delivering genuine feelings of pleasure and happiness. Communications of the ACM, 50 (4). 31-34.

[5] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. and Taylor, J.G. 2001. Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine, 18 (1). 32-80.

[6] Csikszentmihalyi, M. Flow: the psychology of optimal experience. New York: Harper Collins, 1991.

[7] Fairclough, S.H., 2007. Psychophysiological inference and physiological computer games. in Brainplay'07: Brain-Computer Interfaces and Games, Workshop at the Int. Conf. on Advances in Computer Entertainment, (Salzburg, Austria 2007).

[8] Healey, J.A. Wearable and Automotive Systems for Affect Recognition from Physiology Electrical Engineering and Computer Science Dept., MIT, 2000.

[9] Kim, J., 2004. Emotion Recognition from Physiological Measurement. in Humaine European Network of Excellence Workshop, (Santorini, Greece, 2004).

[10] Lang, P.J., Greenwald, M.K., Bradley, M.M. and Hamm, A.O. 1993. Looking at pictures: affective, facial, visceral, and behavioral reactions. Psychophysiology, 30 (3). 261-273.

[11] Mandryk, R.L. and Atkins, M.S. 2007. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. Int. Journal of Human-Computer Studies, 65 (4). 329-347.

[12] Morris, J.D. 1995. SAM:The Self-Assessment Manikin, An Efficient Cross-Cultural Measurement of Emotional Response. Journal of Advertising Research, 35 (6).

[13] Picard, R.W. Affective computing. The MIT press, 1997.

[14] Prensky, M. Computer games and learning : digital game-based learning in Handbook of computer games studies Raessens, J. and Goldstein, J. eds., The MIT Press, Cambridge, MA, 2005.

[15] Rainville, P., Bechara, A., Naqvi, N. and Damasio, A.R. 2006. Basic emotions are associated with distinct patterns of cardiorespiratory activity. International Journal of Psychophysiology, 61 (1). 5-18.

[16] Rani, P., Sarkar, N. and Liu, C., 2005. Maintaining Optimal Challenge in Computer Games through Real-Time Physiological Feedback. in 11th HCI International, (Las Vegas, USA, 2005), Lawrence Erlbaum Associates, Inc.

[17] Sinha, R., Lovallo, W.R. and Parsons, O.A. 1992. Cardiovascular differentiation of emotions. Psychosomatic Medicine, 54 (4). 422-435.