Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty

Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun, *Member, IEEE*

Abstract—This paper proposes to maintain player's engagement by adapting game difficulty according to player's emotions assessed from physiological signals. The validity of this approach was first tested by analyzing the questionnaire responses, electroencephalogram (EEG) signals, and peripheral signals of the players playing a Tetris game at three difficulty levels. This analysis confirms that the different difficulty levels correspond to distinguishable emotions, and that, playing several times at the same difficulty level gives rise to boredom. The next step was to train several classifiers to automatically detect the three emotional classes from EEG and peripheral signals in a player-independent framework. By using either type of signals, the emotional classes were successfully recovered, with EEG having a better accuracy than peripheral signals on short periods of time. After the fusion of the two signal categories, the accuracy raised up to 63%.

Index Terms—Electroencephalography, emotion assessment, games, pattern classification, signal analysis.

I. 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 [1]. Games are also interesting from a human–computer interaction point of view, because they are an ideal ground for the design of new ways to communicate with machines. Affective computing [2] 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, which is to provide emotional experiences such as fun and excitement, affective computing is a promising area of research to enhance game experiences. Affective information can be used to maintain involvement of a player by adapting game difficulty or content to induce particular emotional states [3]. For this purpose, 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/her gaming experience (like it would be the case by using questionnaires). This paper thus focuses on emotion assessment from physiological signals in the context of a computer game application.

Physiological signals can be divided into two categories: those originating from the peripheral nervous system [e.g., heart rate, electromyogram, galvanic skin response (GSR)] and those coming from the central nervous system [e.g., electroencephalogram (EEG)]. In recent years, interesting results have been obtained for emotion assessment with the first category of signals. Very few studies, however, have used the second category, even though the cognitive theory of emotions states that the brain is heavily involved in emotions [4].

One of the pioneering work on emotion assessment from peripheral signals is [5] where the authors detected eight self-induced emotional states with an accuracy of 81%. In [6], six emotional states, elicited by film clips, were classified with an accuracy of 84%. In a gaming context, Rani *et al.* [7] proposed to classify three levels of intensity for different emotions. The emotions were elicited by stimulating participants with a Pong game and anagram puzzles. The best average accuracy obtained with this method was of 86%. The classifiers developed in this paper were used in [3] to adjust game difficulty in real time based on anxiety measures. In this case, the accuracy dropped to 78%, but a significant improvement of player experience was reported compared to difficulty adjustment based on performance. This demonstrates the interest of using affective computing for the purpose of game adaptation. In [8], the authors proposed to continuously assess the emotional state of a player using an approach based on fuzzy logic. The obtained results showed that the emotional state evolved according to the events of the game, but no exact measure of performance was reported. Nevertheless, this tool could be used to include the player's experience in the design of innovative video games. In [9], three emotional states were detected from peripheral signals with an accuracy of 53%. The emotions were elicited by using a Tetris game. This paper is a significant extension of this work, which, in particular, now takes into account the analysis of EEG signals.

There is an increasing amount of psychological literature pointing toward the hypothesis that emotions result from a series of cognitive processes [10], [11]. There is also evidence of different patterns of brain activity during the presentation of emotional stimuli. For instance, depending on the nature of reactions (approach or withdrawal), Davidson [12] showed prefrontal lateralization of alpha waves as well as distinct activations of the amygdala. Aftanas *et al.* [13] reported differences in event-related desynchronization/synchronization during the visualization of more or less arousing images. In the emotional recall context, Smith *et al.* [14] showed an augmentation of

G. Chanel and T. Pun are with the Computer Science Department, University of Geneva, 1227 Carouge, Switzerland (e-mail: guillaume.chanel@unige.ch; thierry.pun@unige.ch).

C. Rebetez and M. Bétrancourt are with the Technologies de Formation et d'Apprentissage Laboratory, Faculty of Psychology, University of Geneva, 1211 Geneva 4, Switzerland (e-mail: cyril.rebetez@unige.ch; mireille.betrancourt@unige.ch).
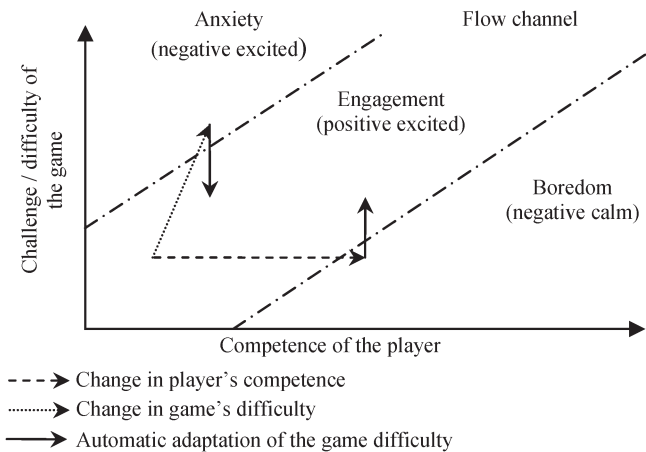
Fig. 1. Flow chart and the suggested automatic adaptation to emotional reactions.

activity in the connections between the hippocampus and the amygdala during the recollection of negative events compared to neutral events. These works emphasize the importance of using brain signals to improve temporal resolution and classification accuracy in emotion assessment. Among the studies that recognize emotional states from EEG, Takahashi [15] obtained an accuracy of 42% to recognize five emotional states elicited by film clips. In [16], three self-induced emotional states were recognized with an accuracy of 68%. Other works tried to infer operator engagement, fatigue, and workload by using EEG signals in order to adapt the complexity of a task [17]–[21]. To our knowledge, however, this paper is the first to report on the use of EEG signals for emotion assessment in a gaming paradigm.

Games can elicit several emotional states, but knowing all of them is not necessary to maintain involvement in the game. Many representations of the player's affective state have been used in previous studies like anxiety, frustration, engagement, distress scales, and the valence-arousal space [22], [23]. According to emotion and flow theories [10], [24], strong involvement in a task occurs when the skills of an individual meet the challenge of a task (Fig. 1). Too much challenge would increase workload which would then be appraised by the player as anxiety. Similarly, not enough challenge would induce boredom. Both these situations would restrain the player's ability to achieve a "flow experience," leading to less involvement, engagement, and possibly interruption of the game [25].

In a game, the change from an emotional state to another can occur due to two main reasons. First, the difficulty is increased because of the progression in different levels, but the increase is too fast compared to the competence increase of the player (potentially giving rise to anxiety; see Fig. 1). Second, the competence of the player has 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 adapt their difficulty according to the competence and emotions of the player. Based on this theory, we defined three emotional states of interest that correspond to three well-separated areas of the valence-arousal space: boredom (negative calm), engagement (positive excited), and anxiety (negative excited).

This paper attempts to verify the validity and usefulness of the three defined emotional states 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 by using the acquisition protocol described in Section II. Using those data, three analyses were conducted. The first aims at validating the applicability of the flow theory for games (see Section III). In the second analysis, detailed in Section IV, physiological signals were used for the purpose of classification of the different states. In this case, since one of the goals of this paper is to go toward applications, particular attention was paid to designing classifiers that could be used for any gamer without having to retrain it.

II. DATA ACQUISITION

A. Acquisition Protocol

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 occurred); and it is playable using only one hand, which is mandatory since the other hand is used for the placement of some data acquisition sensors.

The difficulty levels implemented in the Tetris game were adapted to have a wider range of difficulties than in the original game. The new levels ranged from 1 to 25 with the blocks going down a line every 0.54 s at level 1 and 0.03 s at level 25. The speed of the falling blocks at the intermediate levels increased exponentially with the level. Other modifications to the original Tetris allowed playing without changing the difficulty level for a given amount of time. Each time the blocks reach the top of the Tetris board, a game-over event was reported, the board was cleared, and the participant could continue to play.

Twenty participants (mean age is 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 repeating three times the threshold method, starting from a low level and progressively increasing it until engagement was reported by the participant or starting from a high level and decreasing it. The average of the obtained levels was then considered as the participant's 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 eight levels of difficulty from the player's skill level), and hard condition (higher difficulty, computed by adding eight levels). The participants of the study reported to be engaged at different levels ranging, for most of them, from 11 to 16, confirming that they had different Tetris skills.

Participants were then equipped with several sensors to measure their peripheral physiological activity: a GSR sensor to measure skin resistance, a plethysmograph to record blood volume pulse (BVP), a respiration belt to estimate chest cavity expansion, and a temperature sensor to measure palmar changes in temperature. Those sensors are known to measure signals
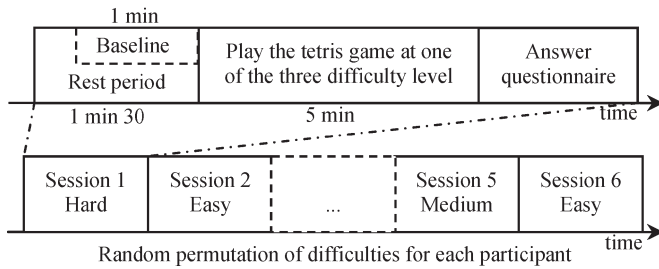
Fig. 2.   Schedule of the protocol.

TABLE I
ENERGY FEATURES COMPUTED FOR EACH ELECTRODE
AND THE ASSOCIATED FREQUENCY BANDS

Feature for electrode i	Frequency band
θ_i	4-8 Hz
α_i	8-12 Hz
β_i	12-30 Hz

that are related to particular emotional activations as well as useful for emotion detection (see Section II-B). In addition, an EEG system was used to record central signaling from 14 of the 20 participants. In this paper, 19 electrodes were positioned on the skull of the participants according to the 10–20 system [26]. As demonstrated in other studies, EEGs can help in assessing emotional states and is also useful in providing an index of task engagement and workload [17]–[20]. Peripheral and EEG signals were recorded at a 256-Hz sampling rate using the Biosemi Active 2 acquisition system.[1] This sampling rate allows keeping the frequency bands of interest for this study.

Once equipped with the sensors, the participants took part in six consecutive sessions (Fig. 2). For each session, the participants had to follow three steps: stay calm and relax for at least 1 min and 30 s, play the Tetris game for 5 min in one of the three experimental conditions (difficulty level), and finally answer a questionnaire. The first step was useful to let the physiological signals return to a baseline level, to record a baseline activity, and to provide a rest period for the participants. For the second step, 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 possible score. To motivate them toward this goal, a prize of 20 Swiss francs was offered to three of the participants having the highest score (the participants were divided in three groups according to their competence). The questionnaire was composed of 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 seven-point Likert scale. Additionally, participants rated their emotions in the valence-arousal space using the self-assessment manikin [27] scales.

B. Feature Extraction

Once the data are acquired, it is necessary to compute features from the signals in order to characterize physiological activity for the different gaming conditions. The features were generally computed over the complete duration of a given session, except in Section IV-D where the features were computed on shorter time windows to analyze the effect of time on emotion-assessment accuracy. Two sets of features were computed: the first set includes the features computed from the EEG signals, and the second includes those computed from the peripheral signals.

In this paper, the collected data are not analyzed for each participant separately but as a whole. It is, thus, necessary

that the patterns of emotional responses remain stable across participants. Although different patterns of emotional responses have been found in psychophysiological studies, Stemmler [28] argues that they are due to context deviation specificity. Since, in the current study, the emotions are elicited in the same context (the video game), this should reduce interparticipant variability. Nevertheless, to further reduce this variability, the physiological signals acquired during the last minute of the rest period were used to compute a baseline activity for each session (six baseline per participant) that was subtracted from the corresponding physiological features.

1) EEG Features: Prior to extracting features from EEG data, we need to remove noise by preprocessing the signals. Environment noise and drifts were removed by applying a 4–45-Hz bandpass filter. The signals were visually checked in order to ensure that the remaining artifacts did not exceed 5% of the signal. The second step was to compute a local reference by applying a local Laplacian filter [29] to render the signals independent of the reference electrode position and to reduce artifact contamination. For the Laplacian filter computation, the neighboring electrodes were considered as lying in a radius of 4 cm from the filtered electrode.

The set of features described in this section was defined to represent the energy of EEG signals in frequency bands known to be related to emotional processes [12], [13]. For each electrode i, the energy in the different frequency bands displayed in Table I was computed for a session, using the fast Fourier transform (FFT) algorithm. Moreover, the following EEG_W feature (1) was computed from the N_e electrodes. This feature is known to be related to cognitive processes like workload, engagement, attention, and fatigue [20], which are cognitive states of interest in our paper. In many studies, the EEG_W feature is computed from only three to four electrodes [17], [18], [20]. However, there is high discrepancy among studies in the electrodes used. Moreover, the playing of a video game can stimulate several brain areas (for instance, the occipital lobe for visual processing, the auditory cortex of the parietal and temporal lobes, and the frontal lobe for emotional processing). For those reasons, all the electrodes were included in the computation of the EEG_W feature

$$EEG\_W = \log\left(\frac{\sum_{i=1}^{N_e} \beta_i}{\sum_{i=1}^{N_e} \theta_i + \alpha_i}\right). \qquad (1)$$

The EEG_FFT feature set thus contains a total of $3 \times 19 + 1 = 58$ features (three frequency bands and 19 electrodes plus the EEG_W feature).

2) Peripheral Features: Many studies in psychophysiology have shown correlations between signals of the peripheral nervous system and emotions; effectiveness of such signals in emotion assessment is now fully demonstrated as detailed in

---

[1]Technical details are available at http://www.biosemi.com.

TABLE II
FEATURES EXTRACTED FROM PERIPHERAL SIGNALS

Peripheral signal	Feature name	Extracted feature	Comment
GSR	μ_{GSR}	Mean skin resistance	Estimate of general arousal level
	δ_{GSR}	Mean of derivative	Average GSR variation
	$f_{GSR}^{DecRate}$	Mean of derivative for negative values only	Average decrease rate during decay time
	$f_{GSR}^{DecTime}$	Proportion of negative samples in the derivative vs. all samples	Importance and duration of the resistance fall
	$f_{GSR}^{NbPeaks}$	Number of resistance falls in the signal	-
BVP	μ_{BVP}	Mean value	Estimate of general BVP
	σ_{BVP}	Standard deviation	BVP variation
Heart rate	μ_{HR}	Mean of heart rate	-
	δ_{HR}	Mean of heart rate derivative	Estimations of heart rate variability
	σ_{HR}	Standard deviation of heart rate	
	f_{HR}^{LF}	Energy in 0.05Hz-0.15Hz band	Parasympathetic and sympathetic activity
	f_{HR}^{HF}	Energy in 0.15Hz-1Hz band	Parasympathetic activity
	$f_{HR}^{LF/HF}$	Ratio of energy in the LF and HF bands	Ratio of parasympathetic and sympathetic activity
Chest cavity expansion	f_{Resp}^{Rate}	Frequency with the highest energy	Respiration rate
	σ_{Resp}	Standard deviation	Variation of the chest cavity expansion signal
	f_{Resp}^{DR}	Maximum value minus minimum value	Dynamic range or greatest breath
Skin Temperature	μ_{Temp}	Mean value	-
	δ_{Temp}	Mean of derivative	Estimation of temperature variability

the introduction. All data were first filtered by a mean filtering to remove noise. For this purpose, we used a rectangular filter of length 128 for GSR, 128 for temperature, and 64 for chest cavity expansion.

GSR provides a measure of the resistance of the skin (electrodermal activity) by positioning two electrodes on the distal phalanges of the index and middle fingers. This resistance decreases due to an increase of sudation, which usually occurs when one is experiencing emotions such as stress or surprise. Moreover, Lang *et al.* discovered that the mean value of the GSR is related to the level of arousal [30]. The number of GSR falls was also computed by identification of the signal local minima. The features extracted from electrodermal activity are presented in Table II.

A plethysmograph was placed on the thumb of the participant to evaluate the BVP. This signal is not only used as a measure of BVP but also to compute heart rate (HR) by identification of local minima (i.e., foot of the systolic upstroke) and interbeat periods. Blood pressure and HR variability are variables that correlate with defensive reactions [31], pleasantness of a stimuli

[30], and basic emotions [32]. The HR signal energy in low frequencies (0.05–0.15 Hz) and high frequencies (0.15–1 Hz), as well as the ratio of these energies, was computed because they are indicators of parasympathetic and sympathetic activities [33].

Chest cavity expansion was measured by tying a respiration belt around the chest of the participant. Slow respiration is linked to relaxation, while irregular rhythm, quick variations, and cessation of respiration correspond to more aroused emotions like anger or fear [32], [34]. To characterize this process, we rely on features from both the frequency and time domain (Table II).

Skin temperature was measured by placing a sensor on the distal phalange of the ring finger. Ekman *et al.* [35] found a significant increase of skin temperature for anger compared to his five other basic emotions (sadness, happiness, fear, surprise, and disgust). McFarland [36] found that stimulating persons with emotional music led to an increase of temperature for calm positive music and a decrease for excited negative pieces.

III. ANALYSIS OF QUESTIONNAIRES AND OF PHYSIOLOGICAL FEATURES

In this section, the data gathered from the questionnaires and from the computed physiological features are analyzed to control the applicability of the flow theory for games. For this purpose, the validity of the following two hypotheses was tested.

1) H1: Playing in the three different conditions (difficulty levels) will give rise to different emotional states.
2) H2: As the skill increases, the player will switch from an engagement state to a boredom state (see Fig. 1).

A. Elicited Emotions

1) Questionnaires: To test for hypothesis H1, a factor analysis was performed on the questionnaires to find the axes of maximum variance. The first two components were obtained from the factor analysis account for 55.6% of the questionnaire variance and were found to be associated with higher eigenvalues than the other components (the eigenvalues of the first three components are 10.2, 8.2, and 1.7). The questionnaire answers given for each session were then projected in the new space formed by the two components, and an analysis of variance (ANOVA) test was applied to those new variables to check for differences in the distribution of judgment for the different conditions. By looking at the weights of the two components, the following was found.

1) The first component was positively correlated with the questions related to pleasure, amusement, interest, and motivation.
2) The second component was positively correlated with the question corresponding to levels of excitation and pressure and negatively correlated with calm and control levels.

The ANOVA test, applied on the data projected on the first component (see Fig. 3), showed that participants felt lower pleasure, amusement, interest, and motivation for the easy and hard conditions than for the medium one ($F = 46, p < 0.01$). Differences in the three distributions obtained from the second
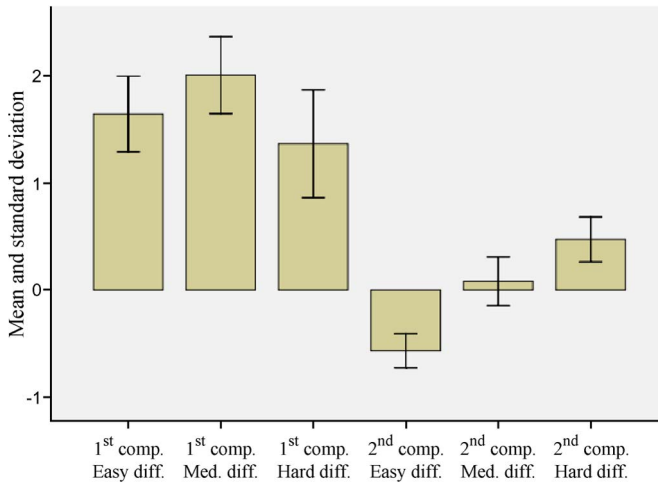
Fig. 3. Mean and standard deviation of judgment for each axis of the two (comp.) component space and the different (diff.) difficulties: easy, (med.) medium, and hard.

component demonstrated that increasing difficulty led to higher reported excitation and pressure as well as lower control ($F = 232, p < 0.01$). This demonstrates that an adequate level of difficulty is necessary to engage players in the game so that they feel motivated and pleased to play. Moreover, those results also validate hypothesis H1 since they show that the different playing difficulties successfully elicited different emotional states with various levels of pleasure and arousal. According to the self-evaluations, those states were defined as boredom for the easy condition, engagement for the medium condition, and anxiety for the hard condition.

2) Peripheral Features: The physiological features were subjected to an ANOVA test to search for differences in activation for the different conditions and analyze the relevance of those features for emotion assessment. For this purpose, the ANOVA test was applied on the three distributions, and the F-values and p-values are reported in Table III. Moreover, the ANOVA test was also applied to check for differences between the easy and medium conditions as well as between the medium and hard conditions. If a difference is significant (p-value < 0.1), the trend of the mean from a condition to another is reported in Table III.

The decrease observed for the μ_{GSR}, δ_{GSR}, and $f_{GSR}^{DecRate}$ features and the increase of the $f_{GSR}^{NbPeaks}$ between the easy and medium conditions indicate an increase of electrodermal activity when progressing from the easy to the medium difficulty level. Between the easy and medium conditions, a significant decrease of temperature is also observed. Those results are in favor of an increase of arousal between the easy and the medium conditions. More specifically, the increase in the number of GSR peaks indicates that the changes in arousal are not only due to workload increase but also to some specific events that triggered emotional reactions. When analyzing the GSR feature changes between the medium condition and the hard conditions, only the $f_{GSR}^{DecTime}$ feature (percentage of negative samples in the GSR derivative) is significantly increasing. An increase of mean HR and a decrease of temperature are also observed between the same conditions. Those results suggest that there is also an increase of arousal between the medium and hard conditions but to a lesser extent than between the easy

TABLE III
F-VALUES AND p-VALUES OF THE ANOVA TESTS APPLIED ON THE PERIPHERAL FEATURES FOR THE THREE DIFFICULTY LEVELS. ONLY THE RELEVANT FEATURES ARE PRESENTED (p-VALUE < 0.1). THE "TREND OF THE MEAN" COLUMN INDICATES THE DIFFERENCES BETWEEN TWO CONDITIONS. FOR INSTANCE, ↘↘ INDICATES A SIGNIFICANT DECREASE OF THE VARIABLE FROM THE EASY TO THE MEDIUM CONDITION (FIRST ↘) AND FROM THE MEDIUM TO THE HARD CONDITION (SECOND ↘), WHILE →↗ INDICATES NO SIGNIFICANT DIFFERENCES BETWEEN THE EASY AND MEDIUM CONDITIONS AND A SIGNIFICANT INCREASE TO THE HARD CONDITION

Feature	F-value	p-value	Trend of the mean
μ_{GSR}	4.4	0.01	↘→
δ_{GSR}	2.7	0.07	↘→
$f_{GSR}^{DecRate}$	3.1	0.05	↘→
$f_{GSR}^{DecTime}$	6.7	< 0.01	→↗
$f_{GSR}^{NbPeaks}$	18.3	< 0.01	↗→
μ_{HR}	3.4	0.04	→↗
f_{HR}^{LF}	2.4	0.09	↘↗
σ_{Resp}	5.8	< 0.01	→↗
μ_{Temp}	9.4	< 0.01	↘↘
δ_{Temp}	10	< 0.01	↘↘

TABLE IV
LIST OF THE RELEVANT EEG FEATURES (p-VALUE < 0.1) GIVEN BY FREQUENCY BAND AND ELECTRODE

	Left electrodes	Midline electrodes	Right electrodes
Theta band	C3, T7, P3, P7, O1	Fz, Cz	F4, C4, T8, O2
Beta band	Fp1, P7, O1	Cz	C4, T8, P8, O2

and medium conditions. In summary, an increased arousal is observed for increasing game difficulty, supporting the results obtained from the analysis of the questionnaires.

As can be seen from Table III, a total of ten features were found to have significantly different distributions among the three difficulties. This suggests that the conditions correspond to different emotional states and demonstrates the interest of those features for later classification of the three conditions. One feature of particular interest is f_{HR}^{LF}, which is the HR energy in low-frequency bands, because it has a lower value for the medium condition than for the two others, showing that this condition can elicit particular peripheral activation. This is also one of the only features that can help distinguish the medium condition from the two others.

3) EEG Features: An ANOVA test was also performed on each EEG feature to test for differences among the three conditions. Table IV gives a list of the EEG features that are relevant (p-value < 0.1). No feature corresponding to the energy in the alpha band was significantly different among the three conditions. However, several features in the theta and beta bands were significantly different, which shows their interest for automatic assessment of the three conditions. To illustrate the EEG activity, we focused on the EEG_W feature since it is a combination of the other features and is known to be related to cognitive processes such as engagement and workload [20].
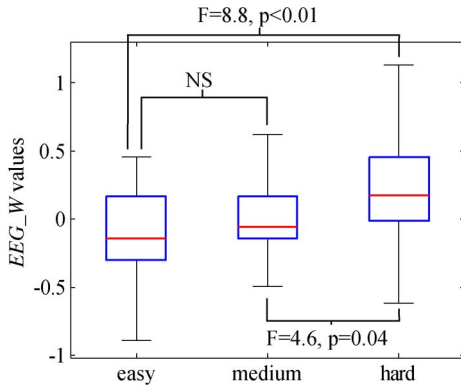
Fig. 4. Boxplot of the EEG_W values for the three conditions. The middle line represents the median of the EEG_W values, the box represents the quartile, and the whiskers represents the range. NS: nonsignificant.

Significant differences were observed for the EEG_W feature among the three conditions ($F = 5.5, p < 0.01$). Fig. 4 shows the median and quartiles of the EEG_W values for each condition. Since for the medium difficulty the participants reported higher interest and motivation than for the easy and hard conditions, it was expected that the mean of the EEG_W values would be significantly higher for the medium condition. However, as can be seen from Fig. 4, there is an increase in the median of the EEG_W values as the difficulty increases. The differences between the medium and hard conditions as well as between the easy and hard conditions are significant according to the ANOVA test. In our view, this reflects the fact that the EEG_W feature is more related to workload than to engagement. The participants involved more executive functions in the hard condition than in the medium one, even if they were less engaged.

B. Evolution of Emotions in Engaged Trials

Hypothesis H2 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 that there was a decrease of engagement from the first session to the second session.

The pairwise t-test used on the variables of the questionnaire showed a significant decrease from the first medium condition to the second medium condition for 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, a decrease in the number of GSR peaks $f_{\mathrm{GSR}}^{\mathrm{NbPeaks}}$ ($t = -2.4, p = 0.02$), as well as an increase in the average of temperature μ_{Temp} ($t = 2.6, p = 0.02$), and in the average of temperature derivative δ_{Temp} ($t = 2.3, p = 0.03$) was found.

Those results are indicative of a decrease of arousal and pleasure while playing twice in the same condition, thus supporting hypothesis H2. The result obtained for the question "I had to adapt to the interface" gives a cue that this decrease could be due to an increase of the player's competence. However, the competence changes were not measured with other indicators to confirm this possibility. In any case, those results demonstrate the importance of having automatic adaptation of the game's difficulty when the challenge of the game remains the same.

IV. CLASSIFICATION OF THE GAMING CONDITIONS USING PHYSIOLOGICAL SIGNALS

A. Classification Methods

In this section, the classification accuracy that can be expected from emotion assessment is investigated. For this purpose, classification methods were applied on the data gathered from the gaming protocol. The ground-truth labels were defined as the three gaming conditions, each one being associated to one of the three states: boredom (easy condition), engagement (medium condition), and anxiety (hard condition).

Three classifiers were applied on this data set: a linear discriminant analysis (LDA), a quadratic discriminant analysis (QDA), and a support vector machine (SVM) with radial basis function (RBF) kernel [37], [38]. The diagonalized versions of the LDA and the QDA were employed because of the low number of samples, which sometimes gives rise to the problem of singular covariance matrices. The size of the RBF kernel was chosen by applying a five-fold cross-validation procedure on the training set and finding the size yielding the best accuracy. The tested size values belonged to the 5.10^{-3}–5.10^{-1} range with a step of 5.10^{-3}.

The following cross-validation method was employed to compute the test accuracy of the classifiers. For each participant, a classifier was trained using the features of other participants; accuracy was then computed by applying the trained model on the physiological data of the tested participant. 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 for a baseline recording of 1 min. Due to the interparticipant variability that remains in physiological activity after baseline subtraction, player-independent classifiers will certainly yield a lower accuracy than player-dependant classifiers. However, this approach allows designing applications where it is not necessary to train a classifier for each user which is drastically time consuming [3].

Three feature-selection algorithms were applied on this problem to find the features that provide good generalization across participants. All those algorithms were applied on the training set to select features of interest, and only the selected features were used for the classification of the test set. An ANOVA feature selection was applied to keep only the features that are relevant to the class concept (*p*-value < 0.1). The fast correlation-based filter (FCBF) [39] was applied to select relevant features and remove redundant ones. The δ_{FCBF} threshold was set to 0.2 because of the following: 1) It was shown in [40] that this value is relevant for FCBF EEG feature selection; and 2) the number of features that has a correlation with the classes higher than 0.2 (7 for peripheral features and 23 for EEG features) is similar to the number of relevant features found using the ANOVA test (10 for peripheral features and 20 for EEG features). Finally, the sequential forward floating selection (SFFS) algorithm [41] was also used to select features of interest, including potentially interacting features. To search for features that have good generalization across participants, the accuracy of a feature subset was estimated by computing the participant cross-validation accuracy on the training
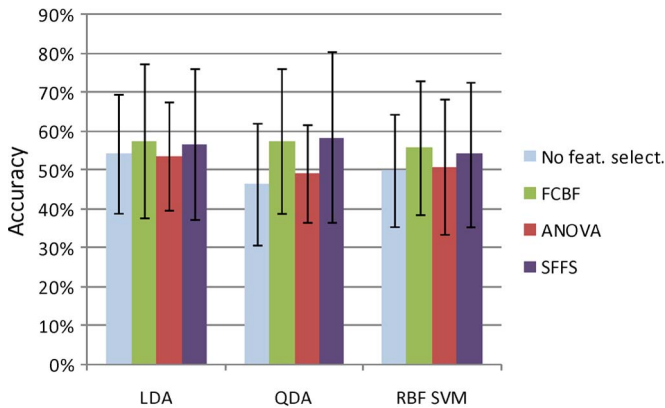
Fig. 5. Accuracies of the different classifiers and feature-selection methods on the peripheral features.

set. The maximum size of a feature subset for the SFFS algorithm was set to 18 for peripheral features and 20 for EEG features.

The fusion of the EEG and peripheral information was performed to improve classification accuracy. This fusion was performed at the decision level [42], by combining the outputs of the classifiers using the Bayes' belief integration [43]. For Bayes' belief integration, the errors produced by the classifiers are expressed by the probabilities $P(y|\hat{y}_q)$ that a classifier q estimates a class as being $\hat{y}_q$, while the true class was y. These probabilities can be computed from the confusion matrices obtained from the training set. The fusion is then performed by assuming classifier independency and choosing the class y that maximizes the following probability:

$$P(y \mid \hat{y}_1 \ldots \hat{y}_{|Q|}) = \frac{\prod\limits_{q \in Q} P(y \mid \hat{y}_q)}{P(y)^{|Q|-1}} \qquad (2)$$

where Q is the ensemble of classifiers used for the fusion.

Since the EEG signals were recorded only for 14 out of the 20 participants, the available number of samples for EEG-based classification is not the same as for peripheral-based classification. For this reason, the results obtained from EEG and peripheral features are separated in two sections with classification algorithm applied on 14 participants for EEG and 20 participants for peripheral features. In Section IV-D, the classification accuracies obtained with EEG and peripheral features on different time scales are compared, while the fusion of peripheral and EEG modalities is investigated in Section IV-E. In both cases, the classification accuracy was computed only on the 14 participants having EEG recorded.

B. Peripheral Signals

Fig. 5 presents the accuracies obtained by applying the classification methods on the features extracted from the peripheral signals. Without feature selection, the LDA obtained the best accuracies of 54% showing its ability to find a boundary that generalizes well across participants. In any case, the accuracies are higher than the random level of 33%. Except for the ANOVA, the feature-selection methods always improved the classification accuracies. The best accuracy of 59% is obtained

with the QDA combined with the SFFS feature selection. However, the FCBF results (58%) are not significantly different from those obtained with the SFFS algorithm because of the high variance of the accuracies. Moreover, the variance of the accuracies obtained with SFFS tends to be higher than those obtained with the FCBF which shows that the FCBF is more stable than the SFFS algorithm in selecting the proper features. According to the results and considering that the FCBF is much faster than the SFFS, the FCBF can be considered as the best feature-selection algorithm for this classification scheme.

Since the participant cross-validation method was used, the feature-selection algorithms were applied 20 times on different training sets. For this reason, the features selected at each iteration of the cross-validation procedure can be different. The histograms of Fig. 6 show, for each feature, the number of times it was selected by a given feature-selection algorithm. The average number of selected features is 3.5 for the FCBF, 9.35 for the ANOVA feature selection, and 4.8 for the SFFS. The ANOVA nearly always selected the features that were found to be relevant in Section III-A but with poor resulting accuracy (Fig. 5). Owing to the removal of redundant features, the FCBF strongly reduces the original size of the feature space with a good resulting accuracy. Moreover, this algorithm nearly always selected the same features independently of the training set showing its stability. The SFFS also obtained good performance, but as can be seen from Fig. 6, some of the features were selected only on some of the training sets, showing that this algorithm is less stable than the FCBF.

By inspecting the SFFS, FCBF, and ANOVA selected features, the $f_{\text{GSR}}^{\text{DecTime}}$ and $f_{\text{GSR}}^{\text{NbPeaks}}$ features were always selected which shows their importance for the classification of the three conditions from physiological signals. To our knowledge, similar features have been used only in [44] for emotion assessment despite of their apparent relevance. The μ_{HR} feature was frequently selected by the FCBF but never by the SFFS and vice versa for the σ_{Resp} feature. The σ_{Resp} feature was removed by the FCBF because it was correlated with μ_{HR}. However, the SFFS kept the σ_{Resp} feature based on its predictive accuracy which suggests that this feature may be better than μ_{HR} for classification. Finally, the temperature features were also found to be frequently relevant.

Because of its good accuracy and low computational time, the FCBF algorithm coupled with QDA classification was used for further analyses involving the peripheral modality. Table V presents the confusion matrix for the three classes: It can be seen that the boredom condition was well classified, followed by the anxiety condition. Samples from the engagement condition tend to be classified mostly as bored samples and also as anxious samples. This is not surprising since this condition lies in between the others. Notice that 21% of the samples belonging to the anxiety class are classified as bored samples; this can be due to the fact that some participants completely disengaged from the task because of its difficulty, reaching an emotional state close to boredom. 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 to correct this problem 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 decrease) to correctly determine the action to take.
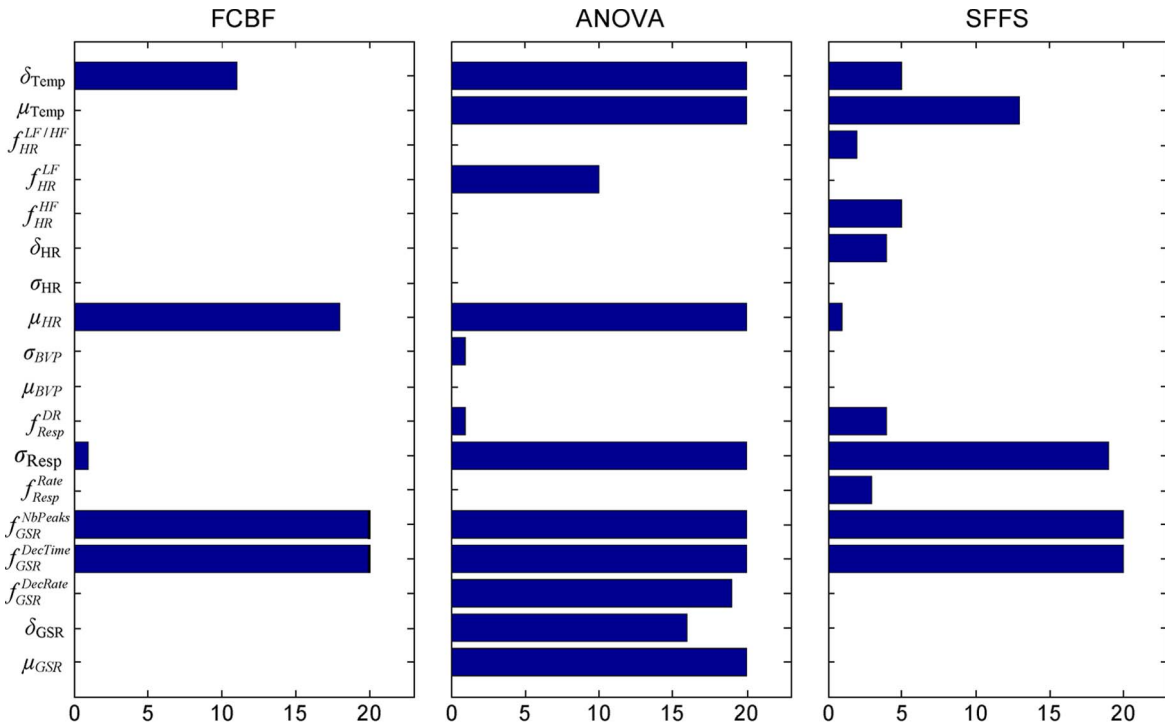
Fig. 6. Histograms of the number of cross-validation iterations (over a total of 20) in which the features have been selected by the FCBF, ANOVA, and SFFS feature-selection algorithms. The SFFS feature selection is displayed for the QDA classification.

TABLE V
CONFUSION MATRIX FOR THE QDA CLASSIFIER
WITH FCBF FEATURE SELECTION

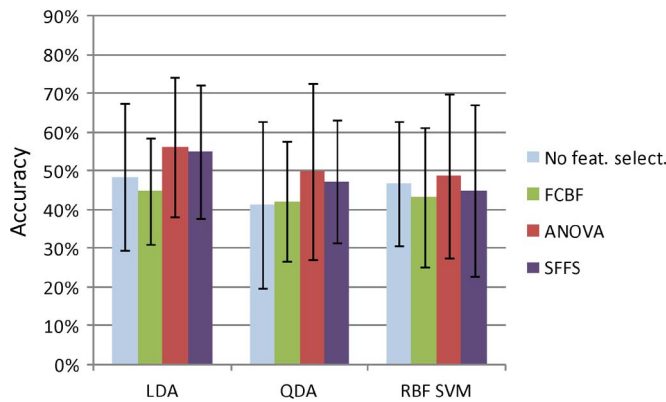| Estimated True	Easy (Boredom)	Medium (Engagement)	Hard (Anxiety)
Easy (Boredom)	80%	10%	10%
Medium (Engag.)	37%	33%	30%
Hard (Anxiety)	21%	19%	60%

Fig. 7. Accuracies of the different classifiers and feature-selection methods on the EEG features.

C. EEG Signals

All the classification methods obtained accuracy higher than the random level of 33% (Fig. 7). Without feature selection, the LDA had the best accuracy of 49%, followed by the RBF SVM with 47%. As with the peripheral features, these results demonstrate the ability of linear and support vector classifiers to well generalize across the participants. The best result of 56% was obtained by the LDA coupled with ANOVA feature selection. The ANOVA feature-selection method always had a

better performance than the other methods. To our knowledge, these are the first results concerning the identification of gaming conditions from EEG signals, particularly considering that the classifiers were trained using a cross-participant framework.

As can be seen from Fig. 8, the FCBF selected less features than the two other feature-selection methods. It selected 3.1 features in average compared to 20.3 for the ANOVA and 13.0 for the SFFS coupled with the LDA. This explains the low accuracy obtained with the FCBF and shows that good accuracies on this problem can be obtained only by concatenating several features. The ANOVA algorithm often selected the features described in Section III-A. The SFFS coupled with the LDA had accuracies close to those of the ANOVA with LDA but by selecting less features in average. For this reason, the features selected by this method are of particular importance for accurate classification of the three gaming conditions. The more often selected features (selected more than eight times) were the theta band energies of the T7, O1, Cz, P4, and P3 electrodes and the beta band energies of the P7, Pz, and O2 electrodes. This result shows that the occipital and parietal lobes were particularly useful for the differentiation of the three gaming conditions.

The confusion matrix displayed in Table VI for the LDA and FCBF methods shows that the different classes were detected with similar accuracies. The medium condition still has the lowest accuracy but is better detected than when using the peripheral features. On the other hand, the easy condition is detected with less accuracy than with peripheral features. This indicates that the fusion of the two modalities should increase the overall accuracy.

D. EEG and Peripheral Signals

In order to compare the accuracies obtained using either EEG or peripheral signals, the best combinations of classifiers and
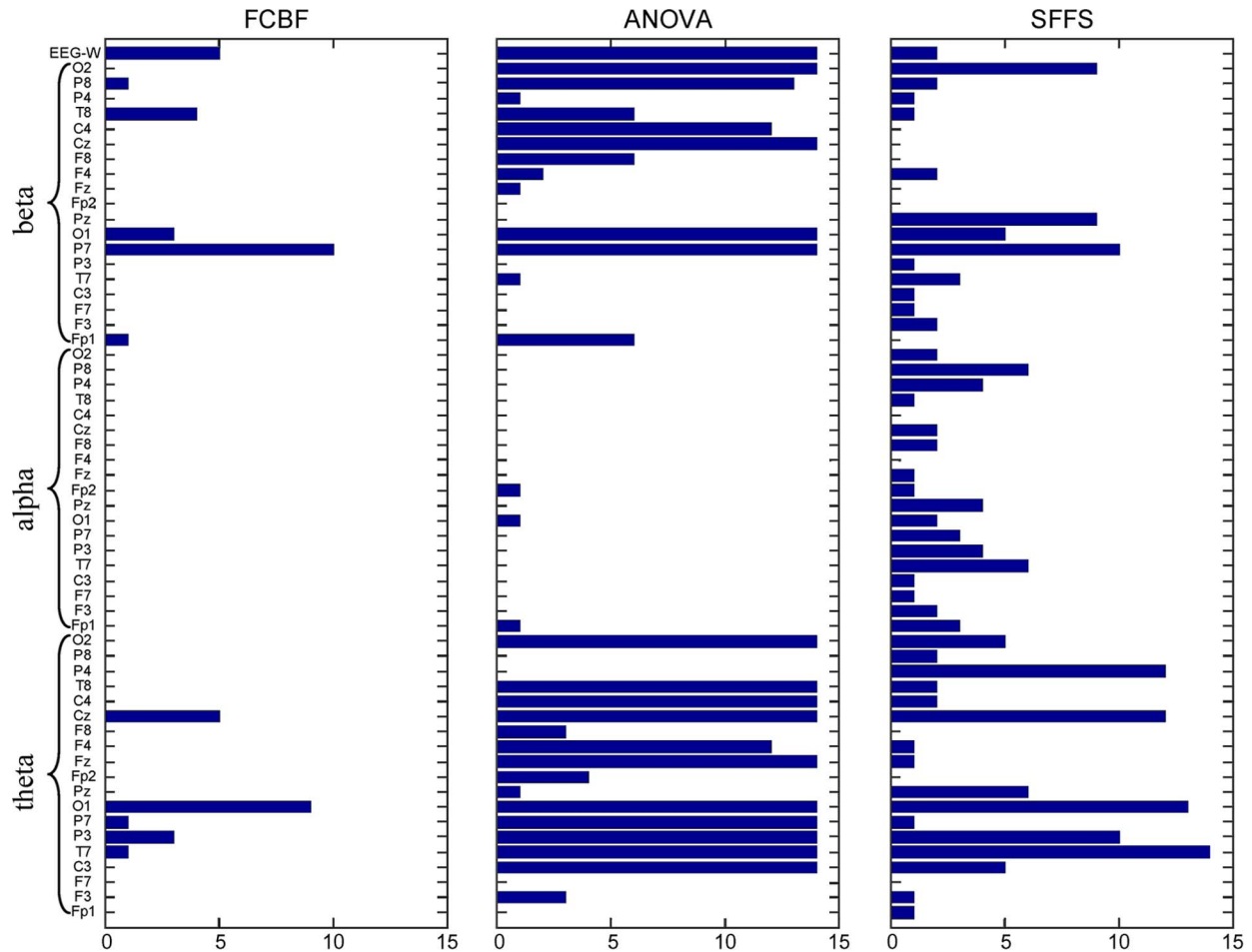
Fig. 8. Histograms of the number of cross-validation iterations (over a total of 14) in which features have been selected by the FCBF, ANOVA, and SFFS feature-selection algorithms. The SFFS feature selection is displayed for the LDA classification.

TABLE VI
CONFUSION MATRIX FOR THE LDA CLASSIFIER
WITH ANOVA FEATURE SELECTION

True \ Estimated	Easy (Boredom)	Medium (Engagement)	Hard (Anxiety)
Easy (Boredom)	57%	43%	0%
Medium (Engag.)	21%	50%	29%
Hard (Anxiety)	19%	19%	62%

feature-selection methods were applied on the physiological database with the same number of participants for both modalities (the 14 participants for whom EEG was recorded). Moreover, the comparison was conducted for different time scales to analyze the performance of each modality as a function of the signal duration used for the feature computation. For this purpose, each session (see Fig. 2) was divided into one to ten nonoverlapping windows of $300/W$ s, where W is the number of windows and 300 s is the duration of a session. EEG and peripheral features were then computed from each window, and the label of the session was attributed to these features. By using this method, a database of physiological features was constructed for each window size ranging from 30 to 300 s.

For a database in which the features were computed from W windows, the number of samples for each class is $20 \times 2 \times W$ (20 participants, 2 sessions per class, and W windows per session). Thus, the number of samples per class increases with W.

Since the number of samples can influence classification accuracy and the goal of this study is to analyze the performance of EEG and peripheral features at different time scales, it is important that this comparison be conducted with the same number of samples for each window's length. To satisfy this constraint, one sample was chosen randomly from each session using a uniform distribution to have $20 \times 2 = 40$ samples per class. The classification algorithms were then applied on this reduced database. This was repeated 1000 times for each value of W to account for the different possible combinations of the windows (except for $W = 1$). Notice that it is not possible to perform classification for all window combinations since there are W^{40} of such combinations.

By using this method, the average accuracies over the 1000 iterations are displayed in Fig. 9. The small accuracy oscillations that can be observed for small time windows (less than 100 s) are likely due to the increase of the number of possible combinations of windows. As can be seen from Fig. 9, the accuracy obtained for the peripheral signals with the original duration of the sessions (300 s) is not significantly different from the one obtained with all of the 20 participants (see Section IV-B). Thus, having 13 or 19 participants for classifier training (because of participant cross validation) does not significantly change the classification performance. This suggests that adding more participants to the current database would not increase classification accuracies, and that, recording
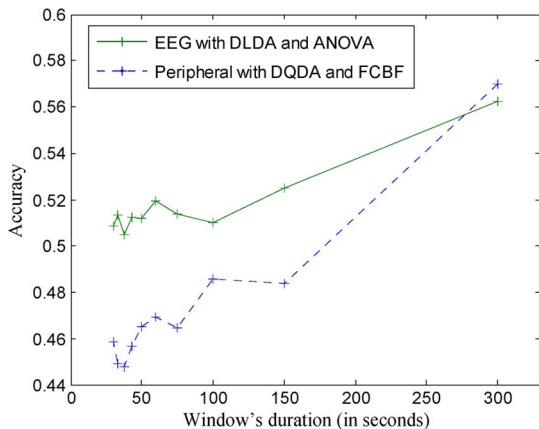
Fig. 9.   Classification accuracy as a function of the duration of a trial for EEG and peripheral features.

TABLE VII
CONFUSION MATRIX FOR THE "BAYES' BELIEF INTEGRATION" FUSION

Estimated True	Easy (Boredom)	Medium (Engagement)	Hard (Anxiety)
Easy (Boredom)	82%	14%	4%
Medium (Engag.)	29%	39%	32%
Hard (Anxiety)	4%	27%	69%

14 to 20 participants is enough to obtain reliable accuracy estimations.

For both modalities, decreasing the duration of the window on which the features are computed leads to a decrease of accuracy. However, this decrease is stronger for peripheral features than for EEG features. For the EEG features, the accuracy drops from 56% for windows of 300 s to around 51% for windows of 30–50 s. For the peripheral features, the accuracy is 57% for windows of 300 s and around 45% for windows of 30–50 s. Moreover, the EEG accuracy remains approximately the same for windows having duration inferior to 100 s, while the peripheral accuracy continues to decrease. All those results demonstrate that the EEG features are more robust on short-term assessment than the peripheral features. For our application, adapting the difficulty of the Tetris game based on the physiological signals gathered during precedent 5 min may be undesirable since there is a high probability that the difficulty of the game has changed during this laps of time due to usual game progress. Having modalities, like EEG, that are able to estimate the state of the user on shorter time periods is thus of great interest.

E. Fusion

As can be seen from the confusion matrices obtained from the classification based on the peripheral and EEG features (Tables V and VI), the errors made with these two feature sets are quite different. The Bayes' belief integration is well suited for this type of problem and, thus, was employed for the fusion of the best classifiers found for each feature set (the LDA couples with ANOVA for EEG features and the QDA couples with FCBF for peripheral features). Another advantage of the Bayes' belief integration is that the probabilities $P(y|\hat{y}_q)$ as in (2) can be estimated independently for the two classifiers. It was thus possible to use the training data of 19 participants

to compute probabilities for the peripheral features, while only 13 participants were used for the EEG features. The resulting accuracy and confusion matrices were obtained by using the participant cross validation applied on the 14 participants for whom both EEG and peripheral activity were recorded.

The accuracy obtained after fusion was 63% which corresponds to an increase of 5% compared to the best accuracy obtained with the peripheral features. Table VII presents the confusion matrix obtained after fusion. By comparing this table to Tables V and VI, it can be observed that the detection accuracy of the easy and the hard classes was increased by 2% and 7%, respectively, compared to the accuracy obtained with the best feature set (peripheral features for the easy class and EEG features for the hard class). The accuracy obtained on the medium class with fusion (39%) is lower than the one obtained with EEG features (50%) but higher than with peripheral features (33%). When performing classification based either on EEG or peripheral features, many of the hard samples were classified as easy, while this problem was solved after fusion. All these results demonstrate the interest of peripheral and EEG fusion at the decision level for a more accurate detection of the three conditions.

The accuracy obtained in the present study is 15% lower than the one obtained in [3]. However, according to the confusion matrix presented in Table VII, the adjusted level of difficulty using the current method should oscillate around the true difficulty level where the participant experiences engagement. It is thus expected that our method will also improve a player's experience. Moreover, as stressed before, the current method only requires a baseline recording of 1 min for each new player, compared to the recording of six 1-h training game sessions for each participant in [3].

V. CONCLUSION

This paper has investigated the possible use of emotion assessment from physiological signals to adapt the difficulty of a game. A protocol has been designed to record physiological activity and gather self-reports of 20 participants playing a Tetris game at three different levels of difficulty. The difficulty levels were determined according to the competence of the players on the task. Two types of analysis have been conducted on the data: First, a statistical analysis of self-reports and physiological data has been performed to control that different cognitive and emotional states were elicited by the protocol; second, classification has been conducted to determine whether it is possible to detect those states from physiological signals.

The results obtained from the analysis of self-reports and physiological data have showed that playing the Tetris game at different levels of difficulty gave rise to different emotional states. The easy difficulty was related to a state of low pleasure, low pressure, low arousal, and low motivation which was determined as boredom. The medium difficulty elicited higher arousal than the easy difficulty, as well as higher pleasure, higher motivation, and higher amusement. It was thus defined as engagement. Finally, the hard condition was associated to anxiety since it elicited high arousal, high pressure, and low pleasure. Moreover, the analysis of consecutive engaged trials has showed that the engagement of a player can decrease if the game difficulty does not change. These results have

demonstrated the importance of adapting the game difficulty according to the emotions of the player in order to maintain his/her engagement.

The classification accuracy of EEG and peripheral signals to recover the three states elicited by the gaming conditions has been analyzed for different classifiers, feature-selection methods, and durations on which the features have been computed. Without feature selection, the best classifiers obtained an accuracy around 55% for peripheral features and 48% for EEG features. The FCBF increased the best accuracy on the peripheral feature to 59%, while the ANOVA selection increased the accuracy to 56% for EEG features. The analysis of the classification accuracy for EEG and peripheral features computed on different duration demonstrated that the EEG features are more robust to a decrease in duration than the peripheral features, which confirms the importance of EEG features for short-term emotion assessment.

Future work will focus on the improvement of the detection accuracy. Fusion of physiological information with other modalities such as facial expressions, speech, and vocal signals would certainly improve the accuracy. Including game information such as the evolution of the score can also help to better detect the three states. Another question of interest is to determine the number of classes to be detected. Since boredom and anxiety are detected with higher confidence than engagement, it might be enough to use those two classes for adaptation to the game difficulty. Moreover, from the observation of Fig. 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 the relations between emotions and competence changes. A future study would be to implement an adaptive Tetris game and verify that it is more fun and enjoyable than the standard one. Finally, analysis of physiological signals for different types of games is also required to see if the results of this study can be extended to other games.

REFERENCES

[1] M. Prensky, "Computer games and learning: Digital game-based learning," in *Handbook of Computer Games Studies*, J. Raessens and J. Goldstein, Eds. Cambridge, MA: MIT Press, 2005.

[2] R. W. Picard, *Affective Computing*. Cambridge, MA: MIT Press, 1997.

[3] C. Liu, P. Agrawal, N. Sarkar, and S. Chen, "Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback," *Int. J. Human-Comput. Interact.*, vol. 25, no. 6, pp. 506–529, Aug. 2009.

[4] K. R. Scherer, *Appraisal Considered as a Process of Multi-Level Sequential Checking*. Oxford, U.K.: Oxford Univ. Press, 2001.

[5] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.

[6] C. L. Lisetti and F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *J. Appl. Signal Process.*, no. 11, pp. 1672–1687, 2004.

[7] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human-robot interaction," *Pattern Anal. Appl.*, vol. 9, no. 1, pp. 58–69, May 2006.

[8] R. L. Mandryk and M. S. Atkins, "A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies," *Int. J. Human-Comput. Stud.*, vol. 65, no. 4, pp. 329–347, Apr. 2007.

[9] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," in *Proc. 12th Int. MindTrek Conf.: Entertainment Media Ubiquitous Era*, 2008, pp. 13–17.

[10] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *IEEE Signal Process. Mag.*, vol. 18, no. 1, pp. 32–80, Jan. 2001.

[11] D. Sander, D. Grandjean, and K. R. Scherer, "A systems approach to appraisal mechanisms in emotion," *Neural Netw.*, vol. 18, no. 4, pp. 317–352, May 2005.

[12] R. J. Davidson, "Affective neuroscience and psychophysiology: Toward a synthesis," *Psychophysiology*, vol. 40, no. 5, pp. 655–665, Sep. 2003.

[13] L. I. Aftanas, N. V. Reva, A. A. Varlamov, S. V. Pavlov, and V. P. Makhnev, "Analysis of evoked EEG synchronization and desynchronization in conditions of emotional activation in humans: Temporal and topographic characteristics," *Neurosci. Behav. Physiol.*, vol. 34, no. 8, pp. 859–867, Oct. 2004.

[14] A. P. R. Smith, K. E. Stephan, M. D. Rugg, and R. J. Dolan, "Task and content modulate amygdala-hippocampal connectivity in emotional retrieval," *Neuron*, vol. 49, no. 4, pp. 631–638, Feb. 2006.

[15] K. Takahashi, "Remarks on emotion recognition from bio-potential signals," in *Proc. 2nd Int. Conf. Auton. Robots Agents*, Palmerston North, New Zealand, 2004.

[16] G. Chanel, J. J. M. Kierkels, M. Soleymani, and T. Pun, "Short-term emotion assessment in a recall paradigm," *Int. J. Human-Comput. Stud.*, vol. 67, no. 8, pp. 607–627, Aug. 2009.

[17] F. G. Freeman, P. J. Mikulka, M. W. Scerbo, and L. Scott, "An evaluation of an adaptive automation system using a cognitive vigilance task," *Biol. Psychol.*, vol. 67, no. 3, pp. 283–297, Nov. 2004.

[18] A. T. Pope, E. H. Bogart, and D. S. Bartolome, "Biocybernetic system evaluates indexes of operator engagement in automated task," *Biol. Psychol.*, vol. 40, no. 1/2, pp. 187–195, May 1995.

[19] M. Besserve, M. Philippe, G. Florence, F. Laurent, L. Garnero, and J. Martinerie, "Prediction of performance level during a cognitive task from ongoing EEG oscillatory activities," *Clin. Neurophysiol.*, vol. 119, no. 4, pp. 897–908, Apr. 2008.

[20] C. Berka, D. J. Levendowski, M. M. Cvetinovic, M. M. Petrovic, G. Davis, M. N. Lumicao, V. T. Zivkovic, M. V. Popovic, and R. Olmstead, "Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset," *Int. J. Human-Comput. Interact.*, vol. 17, no. 2, pp. 151–170, Jun. 2004.

[21] G. F. Wilson and C. A. Russell, "Real-time assessment of mental workload using psychophysiological measures and artificial neural networks," *Hum. Factors*, vol. 45, no. 4, pp. 635–643, Winter 2003.

[22] S. H. Fairclough, "Psychophysiological inference and physiological computer games," in *Proc. Brainplay: Brain-Comput. Interfaces Games, Workshop Int. Conf. Adv. Comput. Entertainment*, 2007.

[23] P. Rani, N. Sarkar, and C. Liu, "Maintaining optimal challenge in computer games through real-time physiological feedback," in *Proc. 11th HCI Int.*, Las Vegas, NV, 2005, pp. 184–192.

[24] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York: Harper Collins, 1991.

[25] K. Salen and E. Zimmerman, *Rules of Play: Game Design Fundamentals*. Cambridge, MA: MIT Press, 2004.

[26] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin. Neurophysiol.*, vol. 112, no. 4, pp. 713–719, Apr. 2001.

[27] J. D. Morris, "SAM: The self-assessment manikin, an efficient cross-cultural measurement of emotional response," *J. Advertising Res.*, vol. 35, no. 6, pp. 63–68, Nov. 1995.

[28] G. Stemmler, M. Heldmann, C. A. Pauls, and T. Scherer, "Constraints for emotion specificity in fear and anger: The context counts," *Psychophysiology*, vol. 38, no. 2, pp. 275–291, Mar. 2001.

[29] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 386–394, Sep. 1997.

[30] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, May 1993.

[31] J. A. Healey, "Wearable and automotive systems for affect recognition from physiology," Ph.D. dissertation, Cambridge, MA, 2000.

[32] P. Rainville, A. Bechara, N. Naqvi, and A. R. Damasio, "Basic emotions are associated with distinct patterns of cardiorespiratory activity," *Int. J. Psychophysiol.*, vol. 61, no. 1, pp. 5–18, Jul. 2006.

[33] G. G. Berntson, J. T. Bigger, Jr., D. L. Eckberg, P. Grossman, P. G. Kaufmann, M. Malik, H. N. Nagaraja, S. W. Porges, J. P. Saul, P. H. Stone, and M. W. van der Molen, "Heart rate variability: Origins, methods, and interpretive caveats," *Psychophysiology*, vol. 34, no. 6, pp. 623–648, Nov. 1997.

[34] J. Kim, "Emotion recognition from physiological measurement," in *Proc. Humaine Eur. Netw. Excellence Workshop*, 2004.

[35] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous-system activity distinguishes among emotions," *Science*, vol. 221, no. 4616, pp. 1208–1210, Sep. 1983.

[36] R. A. McFarland, "Relationship of skin temperature changes to the emotions accompanying music," *Appl. Psychophysiol. Biofeedback*, vol. 10, pp. 255–267, Sep. 1985.

[37] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.

[38] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2001.

[39] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *J. Mach. Learn. Res.*, vol. 5, pp. 1205–1224, 2004.

[40] G. Chanel, K. Ansari-Asl, and T. Pun, "Valence-arousal evaluation using physiological signals in an emotion recall paradigm," in *Proc. IEEE SMC Int. Conf. Syst., Man, Cybern., Smart Cooperative Syst. Cybern.: Adv. Knowl. Security Humanity*, 2007, pp. 2662–2667.

[41] P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proc. IEEE Int. Conf. Pattern Recog.*, 1994, vol. 2, pp. 279–283.

[42] C. Sanderson and K. K. Paliwal, "Identity verification using speech and face information," *Dig. Signal Process.*, vol. 14, no. 5, pp. 449–480, Sep. 2004.

[43] D. Ruta and B. Gabrys, "An overview of classifier fusion methods," *Comput. Inf. Syst.*, vol. 7, no. 1, pp. 1–10, Feb. 2000.

[44] C. D. Katsis, N. Katertsidis, G. Ganiatsas, and D. I. Fotiadis, "Toward emotion recognition in car racing drivers: A biosignal processing approach," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 3, pp. 502–512, May 2008.

Cyril Rebetez received the M.S. degree in learning and teaching technology and the Ph.D. degree in psychology, both from the University of Geneva, Switzerland, with a thesis about multimedia animations for learning, in 2006 and 2009, respectively. The focus of the thesis was to investigate the cognitive processes involved in processing multimedia information and describe the ways to create more understandable and usable media.

At the time of this research, he was with the Technologies de Formation et d'Apprentissage Laboratory, University of Geneva, Switzerland, as a Research and Teaching Assistant in the master of learning and teaching technologies from 2003 to 2009. Since 2010, he has been with Sony Worldwide Studios, London, U.K., as a User Experience Specialist. He is a Researcher interested in human–computer interaction, user experience, video games, multimedia learning, and other related topics.

Mireille Bétrancourt received the M.S. degree in psychology from University of Aix-en-Provence, France, in 1991 and the Ph.D. degree in cognitive sciences from the French National Institute of Technology of Grenoble, France, in 1996.

She is the Head of the Technologies de Formation et d'Apprentissage Laboratory, University of Geneva, Switzerland. She was a Doctoral and Postdoctoral Fellow with the language and Representation Team, French National Institute for Computer Science and Automation. She spent one year as a Postdoctoral Fellow with Stanford University, Stanford, CA. She joined the Faculty of Psychology and Educational Sciences, University of Geneva, in 2000, and was appointed as a Full Professor in information technologies and learning processes in 2003. For over ten years, she has been investigating multimedia learning with two aims: first, providing knowledge about the cognitive processes underlying the comprehension of multimedia and multimodal information; and second, on the basis of cognitive assumptions, investigating how design features affect learning outcomes. Her publication list includes over 60 journal and conference papers.

Guillaume Chanel received the Dipl.Ing. degree in computing from the Institut Méditerranéen d'Etude et de Recherche en Informatique et Robotique, Peripgnan, France, in 2002 and the M.Sc. degree in robotics from the University of Montpellier, France, in 2002. He received his Ph.D. degree in computer science from the University of Geneva, Switzerland, in 2009, where he worked on the automatic assessment of emotions based on electroencephalogram and peripheral signals.

From 2009 to 2010, he was a Researcher with the Knowledge Media Laboratory, Aalto University, Helsinki, Finland. He is currently a Researcher with the Mutimodal Interaction group, Computer Vision and Multimedia Laboratory, Computer Science Department, University of Geneva. His research interests concern the use of physiological measures for improving man–machine interaction and analyzing the mediated social interactions taking place in digital games and serious games.

Thierry Pun (S'79–M'92) received the E.E. Eng. degree and the Ph.D. degree in image processing, for the development of a visual prosthesis for the blind, both from the Swiss Federal Institute of Technology, Lausanne, Switzerland, in 1979 and 1982, respectively.

He is the Head of the Computer Vision and Multimedia Laboratory, Computer Science Department, University of Geneva, Geneva, Switzerland. He was a Visiting Fellow at the National Institutes of Health, Bethesda, MD, from 1982 to 1985. After being a CERN Fellow from 1985 to 1986 in Geneva, he was with the University of Geneva in 1986, where he is currently a Full Professor with the Computer Science Department. He has authored or coauthored about 300 full papers as well as eight patents. His current research interests, related to affective computing and multimodal interaction, concern the following: physiological signal analysis for emotion assessment and brain–computer interaction, multimodal interfaces for blind users, data hiding, and multimedia information retrieval systems.