Physiological-Based Emotion Detection and Recognition in a Video Game Context

Wenlu Yang
Sorbonne University
CNRS, LIP6
Paris, France
wenlu.yang@lip6.fr

Maria Rifqi
Université Panthéon-Assas – Paris 2
LEMMA
Paris, France
maria.rifqi@u-paris2.fr

Christophe Marsala
Sorbonne University
CNRS, LIP6
Paris, France
christophe.marsala@lip6.fr

Andrea Pinna
Sorbonne University
CNRS, LIP6
Paris, France
andrea.pinna@lip6.fr

***Abstract*—Affective gaming is a hot field of research that exploits human emotion for the enhancement of player's experience during gameplay. Physiological signal is an effective modality that can provide a better understanding of the emotional states and is very promising to be applied to affective gaming. Most physiological-based affective gaming applications evaluate player's emotion on an overall game fragment. These approaches fail to capture the emotion change in the dynamic game context. In order to achieve a better understanding of psychophysiological response with a better time sensitivity, we present a study that evaluates the psychophysiological responses related to the game events. More specifically, we present a multi-modal database DAG that contains peripheral physiological signals (ECG, EDA, respiration, EMG, temperature), accelerometer signals, facial and screening recordings as well as player's self-reported event-related emotion assessment through game playing. We then investigate physiological-based emotion detection and recognition by using machine learning techniques. Common challenges for physiological-based affective model such as signal segmentation, feature normalization, relevant features are addressed. We also discuss factors that influence the performance of the affective models.**

I. Introduction

Affective gaming is a hot field of research that exploits human emotion for the enhancement of player's experience during gameplay. To understand the player's emotion, one important research direction for affective gaming is the emotion detection and recognition. Modalities such as facial expression [1], [2], speech [3], [4] have been extensively studied for emotion detection, however, video-based modalities are limited by the lighting and position condition, while audio-based modalities are only applicable to situations where speech exists. On the contrary, using physiological modalities is a very promising approach as it provides continuous, objective, and quantitative measures without the above limitations. Moreover, smart devices such as smart clothes, wristband [5]–[7] or non-contact measurements using video processing technology [8], [9] have been able to non-intrusively measure signals from peripheral nervous system (PNS) (ECG, HR, EDA, etc.) which proves to be a better solution in a practical context. Recent years have seen a lot of studies that explore the interplay between physiological signals and emotion during gameplay [10]. Generally, these studies apply the concepts and methodologies of traditional psychophysiology [11] and contribute to applications such as introducing novice game control [12], [13], game experience augmentation or dynamic difficulty adjustment (DDA) [14]–[16], player modelling [17].

Despite of the interest and advances of affective computing [18] in game [10], only a few databases are publicly available in the affective gaming community. The first public game experience corpus is the Platformer Experience Dataset (PED) [19] that contains game content, behavioural and visual recordings of Super Mario Bros players. The game experience evaluation on engagement, frustration, and challenge are realized on the whole segment of the game. The PED dataset provides a first corpus to model player experience via visual and behavioural cues and self-reported evaluation on the whole game in both ratings and ranks forms. However, the physiological signals are not presented in the dataset and the self-reported evaluations are only realized on the overall game segment which neglects the emotion change related to game events during gameplay. Another dataset is the Mazeball dataset [20], which investigates the effects of camera viewpoints on the psychophysiological state of players by collecting physiological signals (heart rate (HR), blood volume pulse (BVP) and skin conductance (EDA)) and by evaluating the pairwise self-reported emotional preferences in terms of fun, challenge, boredom, frustration, excitement, anxiety, and relaxation. The Mazeball dataset offers the possibility to investigate the physiological features related to player experience. Still, the self-reported assessments are realized at the game segment level. The investigation of the psychophysiological responses related to game events is still unavailable.

In this paper, we first present the multimodal database DAG[1] which contains peripheral physiological modalities (ECG, EDA, Respiration, EMG), behaviour modality (accelerometer signals) as well as the self-reported emotion assessment on both event and segment level. Then, we present a set of analyses aiming at investigating the psychophysiological response using machine learning approaches. Two tasks are considered:

(i) *emotion detection* aims to distinguish the segments with psychophysiological response from those without psychophysiological response. The effects of segmentation lengths and relevant signals are discussed;

[1] http://erag.lip6.fr

(ii) *emotion recognition* aims to distinguish the different emotional states related to the emotional segments. The effects of segmentation lengths and relevant features, as well as three normalization methods (standard normalization, neutral baseline referencing normalization, precedent moment referencing normalization) aiming at reducing individual variabilities are discussed;

The paper is organized as follows: Section II reviews the related work. Section III presents the database and the experience it comes from. Section IV presents the study on emotion detection and recognition. Section V discusses the factors than influence the model's performance. Conclusion and future work are presented in Section VI.

II. Related Work

Physiological-based affective game is an active research topic [10], [21]. Compared with simulations such as image, music and video clips, video game provide a more active and dynamic emotional experience so that dispose several specialities. In this section, we overview the dimensions that characterize the physiological-based affective gaming researches.

1) Game selection: Game selection refers to the selection of game used in the study. In the existing works, there are simple games such as Pong, Pac-Man, Tetris, car racing game [14], [22], [23]. These games only contain a few kinds of events, and are less potential to generate various types of emotions. Meanwhile, there is no direct mapping between the events and emotions. Same event can result in several emotions, which increase research difficulty. There exists games providing more operation liberty such as strategy games or FPS [24], [25]. They offered a richer emotion possibilities and hence represent a better option to analyse physiological response of different emotions during game playing.

2) Modalities: Most physiological modalities presented in the literatures are signals coming from PNS, which confirmes the popularity of using signals from PNS over using central neural system (CNS) in the affective game research. PNS modalities used are modalities related with cardiovascular system such as ECG, HR, BVP, HRV, Respiration, EDA and EMG. Among them, EDA and HR are the most frequently used signals.

3) Representation: The emotion representation largely depends on the genre of game and the research purpose. There are basically 3 types of representation methods:

- representation based on the emotion theory, using dimensional representation such as arousal/valence score [23], [26] or using categorical emotions such as critical emotion in game such as horror, anxiety [22], [24];
- representation based on the flow theory [27] such as boredom, engagement and frustration [14], [28];
- representation based on other dimensions such as preference, attention [29], [30] which represent specific characteristics interested by the researchers.

Representation based on the emotion theory is the most frequently used and also the most flexible. The most frequently used axes are the Russell's original two-dimension (arousal/valence) [31]. Arousal represents the general excitation, ranging from deactivation to activation. Valence means the intrinsic attractiveness/"good"-ness or averseness/"bad"-ness of an event, ranging from unpleasant to pleasant. This method has been widely used in affective computing due to its flexibility [32]–[34]. Concerning the categorical representation, a little difference from the emotion theory is that not all the basic emotions are used in the game research. Selection of the critical emotion is dependent on the type of game. For example, [22] only used "horror", as it is the most frequently occurred emotion in the game they used.

Representation based on the flow theory proposed a high level adaptation objective. The Flow [27] is a balance between the inherent challenge of the game activity and the player's ability to achieve the task. When required skills for the game go beyond players skills, the task becomes too challenging and thus provoks anxiety and frustration. On the contrary, if the task is too simple, the game will fail to engage player and thus evoks boredom. The objective of flow-based adaptation is to try to detect and avoid the frustration or boredom emotion in game to make sure the player stays in the "flow" zone. However, video games often provide complex emotional experiences, so that the high level state of "flow" may depend on a series of low level events and emotions. An overall evaluation based on Flow theory often fails to reveal the mechanism of how "flow" is generated. We can notice that the researchers who adopt this model [14], [28] have only applied adaptation on the simple games such as Tetris or Pac-Man by adapting the speed parameter. When the game becomes more complex, the realization of "flow" state should be based on more detailed low level events and emotions.

4) Time window: Time window refers to the size of time window on which the physiological signals are evaluated for emotion. It represents the temporal sensibility of the emotion recognition. Modalities such EEG or facial expression are reported to have good temporal sensibility [35]. However, the response on PNS are relatively longer. There is still no consensus for the optimal time window for each PNS signal because of the complexity of the physiological signals and the variety of the application contexts.According to the psychophysiological review in [36], the most frequently used time windows are 60, 30 and 10s.

Some researches applied a direct mapping approach rather than an emotion based approach [14], [24]. The direct mapping approach maps the physiological features to a certain game parameter directly and adapt game without recognizing the player's psychological state. The size of time window only controls the frequency of adaptation and can have a wide range [16]. However, this method is not robust as it fails to determine the real psychological state of the user. The emotion-based approach relies more on the appropriate window size. Too short time window may fail to capture the physiological response, while a too long fails to capture the dynamic of emotional experience during the gameplay.

Most emotion-based adaptive affective game research using PNS evaluate the game segment as a whole [15], [22], and

result in long time window and neglect the dynamics of the emotion during the game. As has been put forward by [15], the analysis of physiological signals should also be conducted on the basis of the game events in order to have a better sensitivity.

5) Classification: Some empirical works on emotion measuring explore the direct mapping between the physiological responses and emotions based on expert knowledge in theoretical framework [36], [37], while most application-oriented studies refer to techniques such as machine learning [15], [22], [30] to create data-driven models instead of relying on prior knowledge. As psychophysiological responses in practical contexts are complicated, the major point is that the data-driven approaches may be able to investigate the non-linear multivariate relationship that may otherwise be neglected by the theoretical methods.

III. Database description

We selected a football simulation game FIFA 2016 as emotion stimulation for the following reasons: (i) short repeated events in game can be used to generate different emotions and are easier to get significant analysis; (ii) close relation between emotion and event can be used to provide emotion reference; (iii) changeable difficulty level of the game can be used to provide different experiences. The experiment was conducted at INSEAD - Sorbonne University Multidisciplinary Centre for Behavioural Sciences[2]. In total, 58 participants of different skill levels took part in this study.

A. Modalities and measuring equipments

To analyse physiological responses during game playing, we recorded (i) peripheral physiological signals: ECG, EDA, EMG, respiration and body movement with a 3-axis accelerometer, (ii) facial recording, (iii) game screen recording, (iv) meta-information such as player skill level, game difficulty level, and game resulting score. *Physiological signals* were collected using the BioNomadix wireless sensors and physiology monitoring system Biopac MP150[3]. The sensors used were: an ECG sensor to measure electrocardiogram, an EDA sensor to measure electro-dermal activity, a respiration belt to estimate chest cavity expansion, two ElectroMyoGraphic (EMG) sensors to measure zygomaticus and corrugator muscles movement, an accelerometer sensor to measure body movements. All the used electrodes were Biopac pre-gelled electrodes. In order to reduce the artefacts caused by the movements of hands, the EDA signals have been taken from the arch of the foot.

The game screen output, webcam recording, and screen containing the physiological data were synchronized using a software ObserverXT[4] and visualized on the same screen (Fig. 1) for experimenter.

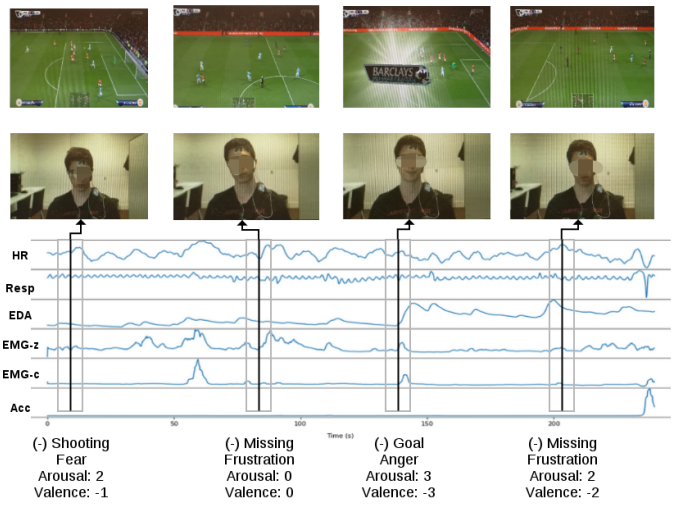


Fig. 1. Experiment scene on one half-time match. The presented elements from top to bottom are: screen recording, player video, physiological signals (HR, Respiration, EDA, pre-processed EMG and pre-processed ACC), and annotations.

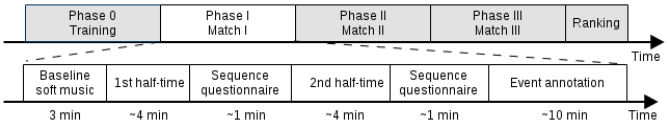


Fig. 2. Protocol of experimentation.

B. Experimental Protocol

Participants played the game in an isolated environment. The procedure of the experiment is presented in Fig. 2. Each experiment was composed of 4 phases: one training phase and 3 match phases. Each match began with 3 minutes of soft music, during which the participant return to neutral state (Fig. 2 Baseline soft music phase). Then, the participant played two half-time (4 minutes for each) of game (Fig. 2 1st half-time and 2nd half-time phase).

After each half-time, the participants filled out a game experience questionnaire (Fig. 2 Sequence questionnaire phase).

At the end of each match, the participant viewed the recording of match and annotated the emotions triggered by significant events during the game (Fig .2 Event annotation phase). To make the annotation easier, we offered participants a list of emotions and events. These lists drawn from the pre-questionnaires during the recruitment process contained the most frequent emotions and events in the game. The elements included in each annotation are:

- events: goal, penalty, shooting, interception of guard, foul, gesture, tackle, corner kick, free kick, arbitration, offside, missing and touch;

[2]http://centres.insead.edu/sorbonne-behavioural-lab/eng/index.cfm
[3]https://www.biopac.com
[4]http://www.noldus.com/the-observer-xt

- emotions: happiness, frustration, proud, curious, angry, fear, boredom, sadness;
- arousal/ valence score: -3, -2, -1, 0, 1, 2, 3.

To speed-up the annotation process, the experimenter helped the participant to annotate with corresponding time-stamp using the software ObserverXT.

In the end, the participants ranked the 3 matches in terms of their perception of difficulty, immersion and amusement (DIA) (Fig. 2 Ranking phase).

IV. Emotion detection and recognition model

The following work lies on the physiological signals and the corresponding self-reported emotion assessments to construct emotion detection and recognition models. We first present the step of feature extraction from a physiological signal segment (Section IV-A), and after, the method and the results for the emotion detection (Section IV-B) and recognition models (Section IV-C). In the end, we show the obtained relevant features and signals for the detection and recognition tasks (Section IV-D).

A. Feature extraction

The general process of feature extraction consists of 3 stages (Table I): (i) *signal pre-processing* cleans the signals to avoid noise or artefacts (such as spiking removing, signal baseline removing, filtering), (ii) *signal transformation* represents the characteristic of a signal in a different aspect (e.g. generating HR sequence from ECG signal), (iii) *feature calculation* extracts common/ specific linear/non-linear time/frequency domain features on the pre-processed or transformed signal. For the *feature calculation*, 2 common feature sets are considered:

1) ***time-domain feature set (time)***

Common time-domain statistic features applied to the sequences are : mean, median, maximum, minimum, range, variance, standard deviation, average derivative, maximum derivative, absolute deviation, kurtosis and skewness. Time–domain feature set contains 12 features in total.

2) ***frequency-domain feature set (freq)***

For feature in frequency domain, as no consensus has been found in the existing works [32], [34], we propose the frequency bands reflecting the main spectral characteristic by covering the principle spectral space. We calculate the band energy of 3 frequency bands in the frequency range [0,4.8] Hz. 4 band energy of the [0,1.6] Hz and the band energy ratios. Frequency-domain feature set contains 8 features in total.

For each segment of signals we obtain 173 features. In the following section, we present the classification tasks and results based on these features.

TABLE I
Feature extraction process

Sig. (nb. fea.)	Preprocess.	Trans.	Feature calcu.
ECG (44)	baseline removing, filtering	raw	raw: *freq*
		IBI	IBI: *time*.
		HRV	HRV: *time.*
		HR	HR: *time*, *freq*.
EDA (53)	spike removing, filtering, subsampling	raw	raw: *time*, *freq*
		phasic	phasic: *time*
		dephasic	dephasic: *time*
		tonic	tonic: *time*
			specific: nb. of peaks
EMG (24)	RMS smoothing, aggregating, subsampling	raw	raw: *time*
Respiration (40)	spike removing, filtering, subsampling, baseline removing	raw	raw: *time*, *freq*
		RR	RR: *time*
		amp.	amp.: *time*
ACC (12)	baseline removeing, RMS smoothing, 3-axis aggregating, subsampling	raw	raw: *time*

raw - pre-processed signals, ***time*** - calculate time domain feature set, ***freq*** - calculate frequency domain feature set

B. Emotion detection model

1) Learning process: Emotion detection distinguishes emotional segments from other segments. By taking advantage of the self-assessed annotations on critical match events, we obtained 2 types of segments for this task: "segments with annotations" which can be considered as emotional moments and "segments without annotation" which can be considered as less likely to be emotional moments.

These 2 types of segments form 2 classes: *emotional* and *non-emotional* class. In order to construct the learning set, we managed to get the same number of instances for each class. For each half-time, we took the segment centred around the annotation points as instances for *emotional* class, and took randomly the same number of non-annotated segments of the same length as instance for the *non-emotional* class.

By taking advantage of the annotation items: *dimensional emotion* and *categorical emotion*, we are interested in detecting each type of emotion. For example, in *dimensional emotion*, *annotated* segments can be categorized into 4 groups: high arousal and high valence (HAHV), high arousal and low valence (HALV), low arousal and high valence (LAHV), low arousal and low valence (LALV). For each type of emotional segments (e.g. HAHV), we took the same number of *non-annotated* segments which represent non-emotional class, in order to construct the learning set to detect (e.g. HAHV).

The learning process is detailed below:

1) ***Segmentation****:* Physiological signals are segmented with different lengths (10s, 14s, 20s, 30s) in order to inspect the most effective signal length to detect the presence of emotions in a dynamic context. The general time-scale to investigate the physiological signals related to affect in based on a whole segment which can sometimes takes

TABLE II
ACCURACY (ACC) AND F1-SCORE (F1) OF EMOTION DETECTION FOR EACH EMOTION GROUP WITH DIFFERENT SEGMENTATION LENGTHS

	10 s			14 s			20 s			30 s		
Groups	**ACC**	**F1**	**Base**	**ACC**	**F1**	**Base**	**ACC**	**F1**	**Base**	**ACC**	**F1**	**Base**
Dimensional Emotion												
HAHV	0.645	0.625**	0.485	0.616	0.629**	0.476	0.630	**0.641****	0.553	0.580	0.604**	0.524
HALV	0.630	0.634**	0.512	0.633	**0.645****	0.523	0.611	0.644**	0.512	0.588	0.644**	0.519
LALV	0.585	**0.515**	0.534	0.570	0.482	0.473	0.552	0.498	0.513	0.495	0.465	0.533
Categorical Emotion												
anger	0.700	**0.689****	0.482	0.688	0.684**	0.498	0.676	0.685**	0.495	0.588	0.597*	0.487
boredom	0.671	0.595*	0.514	0.607	0.583*	0.474	0.533	0.424	0.533	0.640	**0.607***	0.560
fear	0.623	**0.591***	0.530	0.570	0.541	0.473	0.613	0.575*	0.493	0.612	0.574*	0.462
frustration	0.663	**0.685****	0.499	0.635	0.645**	0.502	0.628	0.661**	0.492	0.583	0.629**	0.512
happiness	0.631	0.609**	0.493	0.634	0.624**	0.467	0.619	**0.628****	0.497	0.569	0.595*	0.465

Stars indicate whether the F1-score on detection each type of event is significantly higher than 0.5 according to an independent-samples t-test ($** : p < 0.01, * : p < 0.05$). For comparison, baseline F1-score is given by the maximum between majority and uniform classifier and is presented in the column Base.

as long as several minutes. This setting is evidently not applicable to analyse the psychophysiological response in dynamic context. According to [36], the shortest segmentation length presented is 10s. By varying the segmentation length, we seek the most suitable length. More combinations of different segmentation length on different features and signals can be investigated in the future.

2) ***Feature extraction:*** For each segment, we extract features presented in Section IV-A.
3) ***Normalization:*** In order to reduce the individual variability, each feature is separately normalized for each participant using standard normalization and min-max normalization in the [0,1] range.
4) ***Cross validation:*** We use a 10-fold cross validation scheme.
5) ***Feature selection:*** In each inner loop of the cross validation, we use Fisher's linear discriminant J for feature selection as in [32]:

$$J(f) = \frac{|\mu_+ - \mu_-|}{\sigma_+^2 + \sigma_-^2}$$

where μ_+, μ_- and σ_+, σ_- are the mean and standard deviation of feature f for the positive, negative class respectively. We have tested different sizes of feature set, from 10 to 120 with a step of 10, and finally the best classification rate is obtained for a size of 20.
6) ***Classification:*** Of all the different classifiers (Linear SVM, RBF SVM, Decision tree, Random Forest) we have tested, Linear SVM was the one that achieved the best average accuracy. Hence, we present the results obtained by linear SVM by reporting its accuracy and F1-score. The baseline is taken as the maximum F1-score from the uniform classifier and majority classifier.

2) Detection result: Table II shows the average accuracy of classification of emotional and non-emotional moments.

Among the *dimensional emotion* groups, event groups with high level of arousal (HAHV and HALV) obtain the best result of classification, while the event group with low level of arousal and valence (LALV) obtains the worst result of classification. The performance of event detection on the HAHV and HALV events is significantly better ($p < 0.01$) than on the LALV events. We may conclude that it is easier to detect the HA events than the LA events. Also, whether the event is of high or low valence don't play an important role in the detection efficiency.

Among the *categorical emotion* groups, by observing F1-scores on the 10s segmentation length, event groups with the best detection accuracies are "anger" and "frustration". These emotions are centred on the HALV region of the AV plan which corresponds with our previous observation that events with HALV have the best performance on their detection. "Boredom","fear" and "happiness" detection had a more modest result.

C. Emotion recognition model

1) Learning process: Emotion recognition distinguishes the emotional state for all the emotional segments. This task involves the classification of LA/HA and LV/HV on annotated events ("sequence with annotation"). The ratings of AV on each event is used as learning target. On a scale of 7 points, the AV scores are splitted into two classes: LA/HA classes for arousal classification problem and LV/HV classes for valence classification. Note that the split results in unbalanced classes. To solve this problem, we randomly sample the majority class to get a subset with balanced classes. The classification takes the similar process as in the previous task, except for the ***normalization*** step which is detailed below.

In order to reduce the individual variability and exploit the dynamic of the physiological signals, 3 methods of normalization are applied.

TABLE III
ACCURACY (ACC) AND F1-SCORE (F1) OF EMOTION DETECTION ACROSS PARTICIPANTS

norm.	win(s)	Arousal		Valence	
		ACC	F1	ACC	F1
Std	10	0.477	0.476	0.468	0.502
	14	0.545	0.448	0.546*	0.539
	20	0.532	**0.550***	0.524	**0.567***
	30	0.364	0.362	0.478	0.520
delta	10	0.505	0.509	0.457	0.489
	14	0.559	**0.602****	0.524	**0.573***
	20	0.523	0.534*	0.551*	0.557*
	30	0.477	0.479	0.511	0.526
base.	10	0.508	0.433	0.480	0.461
	14	0.570	**0.551***	0.531*	0.517
	20	0.502	0.473	0.554*	**0.543***
	30	0.453	0.360	0.498	0.473
majority		0,423	0,42	0.455	0,455
uniform		0,504	0,504	0.507	0,507

Stars indicate whether the F1-score on detection each type of event is significantly higher than 0.5 according to an independent-samples t-test ($** : p < 0.01, * : p < 0.05$). For comparison, baseline F1-scores of classification by majority classifier, uniform classifier is presented below.

- *Standard normalization (std)* normalizes each feature for each participants in such a way to have zero mean and unit standard deviation.
- *Normalization referencing precedent segment (delta)* takes the segment just before the annotated segment as reference level. The difference is calculated between annotation segment and the segment before. Then, a standard normalization is applied on this difference for each participant.
- *Normalization referencing baseline segment (base)* takes the neutral state of each participant during music session as reference level. The difference is calculated between annotated segment and music segment. Then a standard normalization is applied on the new features for each participant.

2) Recognition result: Table III shows the average accuracies of binary classification of AV scores with different segmentation lengths and normalization methods across participants.

By comparing the accuracy and F1-scores, we notice that the classification of AV scores is more difficult than the classification of sequences with and without annotation. The classification of valence is more difficult than arousal, as the best F1-score for arousal classification is 0.602 which is better than valence classification F1-score with 0.573.

Concerning the 3 normalization methods, traditional standard normalization (std), precedent sequence referencing normalization (delta) and neutral state referencing normalization (base), the best result for both arousal and valence prediction is obtained by using the precedent sequence referencing method (delta). This method takes the signal and the emotion dynamic into account by assuming that emotion recognition is more linked with the relative feature change than the absolute feature value. For arousal classification, when considering both the ACC and F1 measures, the second best result is obtained by using the referencing neutral state (base) method. The improvement may be explained by the fact that referencing with neutral state reduces the individual variability. However, no clear improvement can be observed on classification of valence.

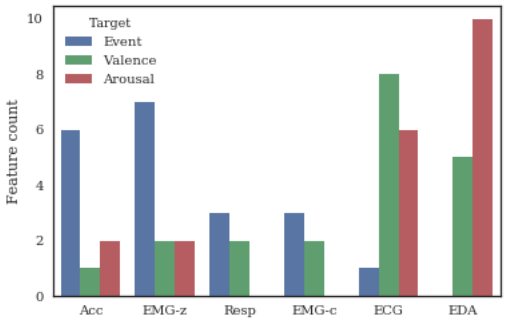


Fig. 3. Importance of modalities for learning tasks: classification of with/without annotation events (*event*), classification of binary arousal score (*arousal*), classification of binary valence score (*valence*)

By comparing the performances of different segmentation lengths, one can notice that the best results are obtained with the segmentation lengths of 14 or 20 seconds, which is longer than the length used in the task of classification of sequences with/without annotation. We may conclude that detection of events requires short segmentation length as longer segmentation smoothes the effects of events, whereas the classification of emotion requires longer sequences as physiological signal varies slowly with emotion, but too long segmentation (e.g. 30s) may cause the overlapping of successive events. As a result, one should find a balance between reaching the necessary signal length for emotion recognition and attaining the optimal time precision to avoid overlapping.

D. Relevant signals for detection and recognition

In order to better understand the role of each signal on the classification results, we select the best 20 features for each task and analyse the most relevant signals for each learning task. Fig. 3 presents the frequency of each signal, from which the features are selected for the emotional event detection and emotion recognition.

One can notice that for emotion detection, the most relevant features are the accelerometer (Acc) and Zygomaticus muscles signals (EMG-z). This further explains why shorter segmentation length is demanded for this task, as the reactions on Acc and EMG-z are instant, longer segmentation length smooth the response effects. For valence classification the most relevant

modalities are ECG and EDA, while for arousal classification, EDA is more important than ECG. This observation is consistent with the work in [38] where ECG is more accurate for classifying valence and EDA for classifying arousal.

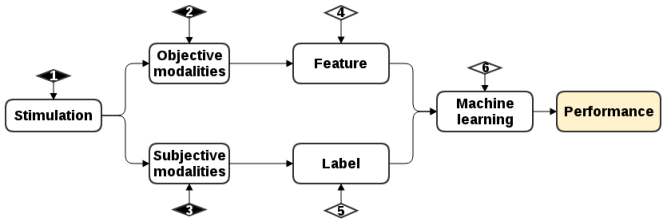


Fig. 4. Process of analysing the affective database.

V. Discussion

Factors that influence the model performance widely exist in each model construction step. Fig. 4 illustrates the process of constructing a machine learning model. In the figure, black points (points 1, 2, 3) present the factors related to the database construction and white points (points 4, 5, 6) present the factors related to the data analysis method.

1) Effectiveness of stimulation: In game context, game scenario varies for each subject, the stimulation cannot be pre-determined, thus making the emotion induction difficult to be controlled. This is an unavoidable choice in analysing self-assessed emotion in dynamic game context, so that the performance may suffer from drawbacks of cognitive error and imprecision of the recalled memory. In future work, efforts can be made to design more controllable game stimulation with a collaborative work of game design company, psychologist, and machine learning practitioner. Moreover, the physiological signals are taken all along the experiment and the timetable of each phase during the experiment (music, gameplay, questionnaire and annotation) is also available. It could be interesting for the researchers who want to perform context detection using physiological signals (pervasive computing [39]).

2) Capacity of objective modalities: The main modality we use in this study is the peripheral physiological signals. In the literature, peripheral physiological signals are mostly used for recognition activation level using long time window (almost 1 min or more) [36]. Psychophysiological responses within short time-window on game event have rarely been addressed [40], [41]. The effectiveness of PPS still needs to be verified by a more detailed study.

3) Effectiveness of subjective modalities: Subjective evaluation is notorious for the noise it may produce due to the cognitive bias. In individual gameplay context, expressive modalities are less presented, so that evaluation by an observer is less evident. The proposed self-reported annotations reflect the participants' emotional state, but suffer from the problems such as cognitive bias, or memory issues. In future work, Besides the self-reported assessment, game event log, the observer/expert's annotation can still serve as a reference.

4) Signal representation - feature: The objective of the feature extraction is to try to reduce the dimension of the input. Different features can be extracted, such as statistical features on the time and frequency domain, entropy-based features, morphological features, or a deep-learning framework can be used to learn a multi-level representation. The choice of the representation should be based on the validated segmentation lengths, as different features of different signals may have different effective length. Besides, the alignment of multi-variant input may also influence the final result, especially in the real-time dynamic context. The present work covers some of the most common features used in the peripheral physiological based affective computing. In future work, more work can be dedicated to find new features extracted from different feature extraction methods or different signal segment lengths.

5) Label processing: Given the cognitive bias which may happen during the subjective evaluation, the obtained labels should be processed in order to reduce this bias. In this paper, techniques such as the discretization of the dimensional evaluation to form a binary classification is applied. In future work, other discretization options, emotion recognition on specific categorical emotions or working directly on dimensional label can be tested.

6) Prediction: A general model cannot achieve good performances on everyone. Due to the complex characteristics of the physiological signals and the subjectiveness of the self-reported assessment, there exists great variability among individuals. In future work, more attention should be paid to individual differences from the model view, for example by investigating the similar subjects and creating models for them.

VI. Conclusions

This paper investigates the psychophysiological response to events on an affective database in a game context. We address some common challenges for physiological-based affective model such as segmentation, normalization, or relevant signals. For emotion detection task, we show that high arousal emotions are more detectable than the low arousal ones. The most relevant features for this tasks derive from the ACC and EMG signals. Concerning the emotion recognition task, we show that it is more difficult than emotion detection, the best result is obtain with a segmentation length of 14 or 20s. The most relevant features derive from EDA, ECG. In the end, we also discuss the factors that influence the performance of the affective models and the future works.

Acknowledgments

This work was performed within the Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02. The experiment has been done thanks to the fund "Soutien au démarrage d'études comportementales" granted by the Centre Multidisciplinaire des Sciences Comportementales, Sorbonne Universités-INSEAD.

Furthermore, the authors gratefully thank Victor Billot for aiding carrying out the experiment and also members of INSEAD-Sorbonne Universités Behavioural Lab: Liselott Pettersson, Jean-Yves Mariette, Germain Dépetasse, Hoai Huong Ngo.

REFERENCES

[1] E. G. Krumhuber, A. Kappas, and A. S. Manstead, "Effects of dynamic aspects of facial expressions: a review," *Emotion Review*, vol. 5, no. 1, pp. 41–46, 2013.

[2] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, and K. Scherer, "Toward a minimal representation of affective gestures," *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 106–118, 2011.

[3] F. Dellaert, T. Polzin, and A. Waibel, "Recognizing emotion in speech," in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, vol. 3. IEEE, 1996, pp. 1970–1973.

[4] V. A. Petrushin, "Emotion recognition in speech signal: experimental study, development, and application," *studies*, vol. 3, no. 4, 2000.

[5] T. Christy and L. I. Kuncheva, "Amber shark-fin: An unobtrusive affective mouse," in *ACHI 2013, The Sixth Intl. Conf. on Advances in Computer-Human Interactions*, 2013, pp. 488–495.

[6] A. Bonarini, F. Costa, M. Garbarino, M. Matteucci, M. Romero, and S. Tognetti, "Affective videogames: the problem of wearability and comfort," in *Human-Computer Interaction. Users and Applications*. Springer, 2011, pp. 649–658.

[7] N. Oliver and F. Flores-Mangas, "Healthgear: a real-time wearable system for monitoring and analyzing physiological signals," in *Wearable and Implantable Body Sensor Networks, 2006. BSN 2006. Intl. Workshop on*. IEEE, 2006, pp. 4–pp.

[8] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcama non-contact method for evaluating cardiac activity," in *Computer Science and Information Systems (FedCSIS), 2011 Federated Conf. on*. IEEE, 2011, pp. 405–410.

[9] K. S. Tan, R. Saatchi, H. Elphick, and D. Burke, "Real-time vision based respiration monitoring system," in *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th Intl. Symposium on*. IEEE, 2010, pp. 770–774.

[10] B. Bontchev, "Adaptation in affective video games: A literature review," *Cybernetics and Information Technologies*, vol. 16, no. 3, pp. 3–34, 2016.

[11] S. H. Fairclough, "Psychophysiological inference and physiological computer games," in *BRAINPLAY 07 Brain-Computer Interfaces and Games Workshop at ACE (Advances in Computer Entertainment)*, 2007, p. 19.

[12] S. I. Hjelm and C. Browall, "Brainball-using brain activity for cool competition," in *Proceedings of NordiCHI*, vol. 7, no. 9, 2000.

[13] J. Sharry, M. McDermott, and J. Condron, "Relax to win treating children with anxiety problems with a biofeedback video game," *Eisteach*, vol. 2, pp. 22–26, 2003.

[14] T. Tijs, D. Brokken, and W. IJsselsteijn, "Creating an emotionally adaptive game," in *International Conference on Entertainment Computing*. Springer, 2008, pp. 122–133.

[15] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," in *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*. ACM, 2008, pp. 13–17.

[16] Z. O. Toups, R. Graeber, A. Kerne, L. Tassinary, S. Berry, K. Overby, and M. Johnson, "A design for using physiological signals to affect team game play," *Foundations of Augmented Cognition*, pp. 134–139, 2006.

[17] P. A. Nogueira, R. Aguiar, R. A. Rodrigues, E. C. Oliveira, and L. Nacke, "Fuzzy affective player models: A physiology-based hierarchical clustering method." in *AIIDE*, 2014.

[18] R. W. Picard *et al.*, "Affective computing," 1995.

[19] K. Karpouzis, G. N. Yannakakis, N. Shaker, and S. Asteriadis, "The platformer experience dataset," in *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*. IEEE, 2015, pp. 712–718.

[20] G. N. Yannakakis, H. P. Martínez, and A. Jhala, "Towards affective camera control in games," *User Modeling and User-Adapted Interaction*, vol. 20, no. 4, pp. 313–340, 2010.

[21] S. H. Fairclough, "Fundamentals of physiological computing," *Interacting with computers*, vol. 21, no. 1-2, pp. 133–145, 2008.

[22] C. Liu, P. Agrawal, N. Sarkar, and S. Chen, "Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback," *International Journal of Human-Computer Interaction*, vol. 25, no. 6, pp. 506–529, 2009.

[23] A. Parnandi and R. Gutierrez-Osuna, "A comparative study of game mechanics and control laws for an adaptive physiological game," *Journal on Multimodal User Interfaces*, vol. 9, no. 1, pp. 31–42, 2015.

[24] A. Dekker and E. Champion, "Please biofeed the zombies: Enhancing the gameplay and display of a horror game using biofeedback." in *DiGRA Conference*, 2007.

[25] L. E. Nacke, M. Kalyn, C. Lough, and R. L. Mandryk, "Biofeedback game design: using direct and indirect physiological control to enhance game interaction," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2011, pp. 103–112.

[26] P. A. Nogueira, R. Rodrigues, and E. Oliveira, "Real-time psychophysiological emotional state estimation in digital gameplay scenarios," in *International Conference on Engineering Applications of Neural Networks*. Springer, 2013, pp. 243–252.

[27] M. Csikszentmihalyi, "Das flow-erlebnis," *Jenseits von Angst und Langeweile: im Tun aufgehen. Stuttgart*, 1985.

[28] S. Fairclough and K. Gilleade, "Construction of the biocybernetic loop: a case study," in *Proceedings of the 14th ACM international conference on Multimodal interaction*. ACM, 2012, pp. 571–578.

[29] L.-D. Liao, C.-Y. Chen, I.-J. Wang, S.-F. Chen, S.-Y. Li, B.-W. Chen, J.-Y. Chang, and C.-T. Lin, "Gaming control using a wearable and wireless eeg-based brain-computer interface device with novel dry foam-based sensors," *Journal of neuroengineering and rehabilitation*, vol. 9, no. 1, p. 5, 2012.

[30] S. Tognetti, M. Garbarino, A. Bonarini, and M. Matteucci, "Modeling enjoyment preference from physiological responses in a car racing game," in *Computational Intelligence and Games (CIG), 2010 IEEE Symposium on*. IEEE, 2010, pp. 321–328.

[31] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, p. 1161, 1980.

[32] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis ;using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, pp. 18–31, Jan 2012.

[33] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, pp. 42–55, Jan 2012.

[34] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe, "Decaf: Meg-based multimodal database for decoding affective physiological responses," *IEEE Transactions on Affective Computing*, vol. 6, pp. 209–222, July 2015.

[35] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, "Introducing the recola multimodal corpus of remote collaborative and affective interactions," in *2013 10th IEEE Intl. Conf. and Workshops on Automatic Face and Gesture Recognition (FG)*, April 2013, pp. 1–8.

[36] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biological psychology*, vol. 84, no. 3, pp. 394–421, 2010.

[37] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition and emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[38] F. Ringeval, F. Eyben, E. Kroupi, A. Yuce, J.-P. Thiran, T. Ebrahimi, D. Lalanne, and B. Schuller, "Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data," *Pattern Recogn. Lett.*, vol. 66, pp. 22–30, Nov. 2015.

[39] J. Ye, S. Dobson, and S. McKeever, "Situation identification techniques in pervasive computing: A review," *Pervasive and mobile computing*, vol. 8, no. 1, pp. 36–66, 2012.

[40] N. Ravaja, M. Turpeinen, T. Saari, S. Puttonen, and L. Keltikangas-Järvinen, "The psychophysiology of james bond: Phasic emotional responses to violent video game events." *Emotion*, vol. 8, no. 1, p. 114, 2008.

[41] N. Ravaja, T. Saari, M. Salminen, J. Laarni, and K. Kallinen, "Phasic emotional reactions to video game events: A psychophysiological investigation," *Media Psychology*, vol. 8, no. 4, pp. 343–367, 2006.