Analysing the Relevance of Experience Partitions to the Prediction of Players' Self-Reports of Affect

Héctor P. Martínez and Georgios N. Yannakakis

Center for Computer Games Research,
IT University of Copenhagen,
Rued Langgaards vej 7, 2300, Denmark
`{hpma,yannakakis}@itu.dk`

Abstract. A common practice in modeling affect from physiological signals consists of reducing the signals to a set of statistical features that feed predictors of self-reported emotions. This paper analyses the impact of various time-windows, used for the extraction of physiological features, to the accuracy of affective models of players in a simple 3D game. Results show that the signals recorded in the central part of a short gaming experience contain more relevant information to the prediction of positive affective states than the starting and ending parts while the relevant information to predict anxiety and frustration appear not to be localized in a specific time interval but rather dependent on particular game stimuli.

Keywords: preference learning, post-experience self-reports, heart rate, skin conductance, blood volume pulse, games

1 Introduction

Video games, even the most simple ones, have the potential of providing engaging episodes in which the player might experience a plethora of psychological states from fear and frustration to excitement and fun. A game able to recognize those psychological states could offer tailored and optimized experiences to each different player according to her preferences and motivations to play without the need of asking directly [1, 2].

A predictor of affect can be trained on data gathered across several game sessions in which players report their experience during or after the game ([3–5] among others) resulting in a computational model that receives as inputs the objective (measurable) part of the experience (e.g. achievements on the game, buttons pressed on the game pad, heart rate and facial expression) and outputs a value estimating the subjective part (e.g. valence and arousal or frustration). Typically, the inputs of the model consist of *statistical features* (e.g. score, keys pressed per second and average heart rate) calculated on a time interval before the player responds to online (i.e. during play) or post-experience questionnaire [6, 7]. According to some theories, when humans report past emotional experiences, they have to retrieve specific thoughts, event-specific details or beliefs [8] that relate to the past experience. Consequently, we expect that certain parts of the experience are more relevant than others to predict self-reported affect. To the best of our knowledge there exists no study that suggests an optimal time interval that maximizes

the predictability of computational models for recognition of affect. The common practice followed is to either choose an arbitrary time window (e.g. the complete game [5]) or run preliminary experiments to determine the most appropriate one for the task at hand [9].

This paper examines the relevance of different experience time windows with respect to the prediction accuracy of seven self-reported affective/cognitive preferences — players report after a pair of games whether the first or the second felt more *anxious*, *boring*, *challenging*, *exciting*, *frustrating*, *fun* and *relaxing* — in a short (90 seconds) 3D prey/predator game. Features calculated on different time intervals of three physiological signals, namely heart rate (HR), skin conductance (SC) and blood volume pulse (BVP), are compared as inputs to artificial neural network (ANN) models of affect.

The rest of the paper is organized as follows: Section 2 reviews the literature in psychophysiology, games and affect, Section 3 and 4 , respectively, present the dataset and the methodology followed in this study and Section 5 and 6 discuss the results and conclusions of the study.

2   Related Work

Research in game psychophysiology has gained interest in recent years [10] resulting in studies exploring different modalities, emotions and affective states of players. Some affective gaming researchers attempt to draw the mappings between physiology and affect by analysing the correlations between statistical features extracted from the physiological signals of participants while playing and their affective post-experience self-reports. For example, Nacke and Lindley [11] investigate the correlations between flow, boredom and immersion, and the cumulative averages of jaw electromyography (EMG) and SC over the complete playing time in a first person shooter while Rani et al. [12] explore the correlations between anxiety, engagement, boredom and frustration self-reports after playing Pong and various statistical features extracted from the physiological signals (HR and SC among others) recorded during the entire game.

Another approach to the same problem consists of training computational models that predict self-reports of affect relying on features of the physiological signals. Kapoor et al. [9] implement different computational predictors of frustration based on the average value over 150 seconds of 14 features extracted from a SC sensor, a pressure mouse, a posture sensor and a face tracker. The time window is calculated just before the participant reports to be frustrated or 225 seconds after the game has started if the participant does not report frustration during the experience. McQuiggan et al. [4] aim to maximize the time for a correct early detection of frustration based on BVP and SC features (among other modalities). The best models reported predict the frustration self-reports 35 seconds before they are introduced by the user. Mandryk et al. [13] use fuzzy rules to map HR, SC, respiration and EMG of the jaw muscles continuously (i.e. every physiological recording is transformed) to an arousal-valence space and further to levels of fun, boredom, challenge, excitement and frustration during a hockey computer game. This model is validated by analysing the correlation between the average value of the predicted psychological states along the entire game session and the post-experience emotional ratings of the players. Yannakakis et al. [14, 15] model the fun

pairwise preferences of children playing physical interactive games from an extensive set of statistical features extracted from HR and SC recorded during the 90-second long games using neuroevolutionary preference learning. On the same basis, Martinez et al. [16, 5] trained predictors of seven self-reported affective states in a 3D prey/predator game using statistical features of HR, BVP and SC recorded during 90-second long games. Tognetti et al. [17] apply Linear Discriminant Analysis to map a large number of features calculated on the last minute recordings of HR, BVP, respiration and SC to player preferences in a racing game.

In the aforementioned studies, either the complete signals over the full length of a game or an interval selected after preliminary — not reported — studies are used to calculate the statistical features. On the contrary, McQuiggan et al. [18] explore different time window lengths and positions in HR and SC signals to calculate the statistical features that feed a predictor of self-efficacy in a 3D learning environment. A large number of features are calculated covering overlapping parts of the experience. Unfortunately, the relevance of each feature for the prediction is not analysed. Broekens et al. [19] present a brief analysis of five window lengths in eye-gaze data. That analysis shows that features calculated in shorter windows (1, 2, 4 and 6 seconds) prior a player action predict more accurately the type of action than longer time windows (10 seconds). That study differs from this paper both in the signals analysed and more importantly the prediction target.

It is worth mentioning other psycho-physiological studies in games with a focus on short signal intervals associated to game events. Conati and Maclaren [20] propose a probabilistic model of joy and distress of students and their admiration and reproach towards an agent based on Dynamic Decision Networks. For each game event, the model takes as input the difference between the average EMG signal in the corrugator muscle over the whole experience (30 minutes) and the average EMG in the four seconds following the event. Hazlett [21] studies the correlations between positive and negative events and the mean value of corrugator and zygomaticus muscles EMG during those events and one second following the event. Ravaja et al. [22] examine the effect of game events on zygomatic and orbicularis oculi EMG, SC and HR by analysing statistically the changes on the mean value of the signals in eight 1-second windows (2 before the event and 6 after). On these studies, the time windows are selected to allow enough time to detect a physiological response while minimizing overlapping windows among subsequent events. This paper does not explore event-associated time windows and instead focuses on models of post-experience affective reports that are based on physiological signals gathered during play.

3   Data Recording

The dataset used in this paper was gathered during an experimental game survey in which 36 participants (80% males, aged from 21 to 47 with mean and standard deviation of age equal 27.2 and 5.84, respectively) played four pairs of different variants of the same video-game. The test-bed game named *Maze-Ball* is a 3D prey/predator game that features a green ball inside a maze controlled by the arrow keys. The goal of the player is to maximize her score in 90 seconds by collecting the pellets scattered in the maze

while avoiding the red enemies that wander around. The eight available variants of the game differ only on the virtual camera profile used which defines how the virtual world is presented on screen.

Blood volume pulse, b, and skin conductance, s, were recorded at 32Hz during the session using the IOM biofeedback device. Heart rate, h, is inferred by the BVP signal every 5 seconds and the magnitude (SM), m, and the duration (SD), d, of signal variation have been derived from SC [23]. The players filled in a 4-alternative forced choice questionnaire after completing a pair of variants reporting whether the first or the second game of the pair felt more anxious, boring, challenging, exciting, frustrating, fun and relaxing, or whether both felt equally, or none of them did. The details of the Maze-Ball game design and the experimental protocol followed can be found in [24, 5].

The following set of features is extracted from different time intervals — full game (90 seconds), two overlapping windows of 60 seconds (0 to 60 and 30 to 90), two non-overlapping windows of 45 seconds, three non-overlapping windows of 30 seconds and four non-overlapping windows of 22.5 seconds — of each signal ($\alpha \in \{b, s, h, m, d\}$) inspired by previous studies on physiological feature extraction [15, 25]:

- average ($E\{\alpha\}$) and variance ($\sigma^2\{\alpha\}$) of the signal;
- initial (α_{in}) and final (α_{last}) recording and difference between them (Δ^α);
- minimum ($min\{\alpha\}$) and maximum ($max\{\alpha\}$) signal recording and difference between them (D^α);
- time when maximum ($t_{max}\{\alpha\}$) and minimum ($t_{min}\{\alpha\}$) samples were recorded and difference between those times (D_t^α);
- average first and second absolute differences ($\delta_{|1|}^\alpha$ and $\delta_{|2|}^\alpha$, respectively);
- Pearson's correlation coefficient (R_α) between raw α recordings and the time t at which data were recorded;
- autocorrelation (lag equals the sampling rate of α) of the signal (ρ^α);

All features are normalized to the [0,1] interval using standard min-max normalization.

4 Method

Neuroevolutionary preference learning is applied in order to train computational models that learn the players' pairwise self-reports of affect. The inputs to the models are selected automatically through Genetic Feature Selection (GFS). The two algorithms are briefly described in the following subsections.

4.1 Genetic Feature Selection

Feature selection (FS) is essential in scenarios where the available features do not have a clear relationship and, thus, impact to the prediction of a target output (i.e. it is not easy to decide *a priori* which features are useful and which are irrelevant for the prediction). Moreover the computational cost of testing all available feature sets is combinatorial and exhaustive search might not be computationally feasible in large feature sets. Under

these conditions, FS is critical for finding an appropriate set of model input features that can yield highly accurate predictors [5].

Genetic feature selection [16] is a global search FS algorithm guided by a genetic search. The search starts by evaluating the fitness of several subsets with one feature; in subsequent iterations combinations of the fittest subsets from the previous iterations are evaluated. The algorithm stops after a fixed number of iterations or when highly fit feature subsets are found. The fitness function is calculated as the average cross validation performance of a model trained on the selected features on unseen folds of classification data. More details about GSF can be found in [16].

4.2 Neuroevolutionary Preference Learning

We apply preference learning [26] to build affective models that predict users' self-reported emotional preferences based on the subsets of features selected by the GFS algorithm. In this study, the models are implemented as single layer perceptrons (SLPs) that are trained via neuroevolutionary preference learning (as in [27, 5]) to map the selected features to a predictor of the reported pairwise emotional preferences.

Note that the pairwise preference relationship of the training data is known (e.g. game A is preferred to game B) but the value of the target output is not (i.e. the magnitude of the preference is unknown). Thus, any gradient-based optimization algorithm is inapplicable to the training problem since the error function under optimization is not differentiable. The trained model learns for each pair of games a higher output value for the preferred game than for the non preferred.

5 Experiments

Ten ANN-models are trained for each affective state and time window using 3-fold cross validation and their average accuracy is depicted on Figure 1. For the 45 and 60 second long intervals, all affective states but challenge are predicted with higher accuracy in the second interval. By further subdividing the windows, boredom and the three positive affective states — excitement, fun and relaxation — are predicted with the highest accuracy using statistical features of the physiological signals calculated in the central 30-second interval and the second or third quarter of the game (22.5 second long intervals). Furthermore, these models built on small time intervals yield, on average, significantly higher prediction accuracies than the models trained on the full-length experience — difference in accuracy of 3.15% for boredom ($t(18) = 3.43$), 6.0% for excitement ($t(18) = 7.31$), 4.75% for fun ($t(18) = 4.52$) and 7.22% for relaxation ($t(18) = 10.82$); *p-values* < 0.01 for all four states. This suggests that the information more relevant for predicting self-reports of positive experiences and boredom in Maze-Ball is located at time windows which are in the middle of the short (90 s) gaming experience.

The anxiety models built on 45 and 60 second-long windows suggest that the second half of the game is more relevant to define the experience; however, the smaller windows do not show a clear trend. Similarly, frustration models present the same average
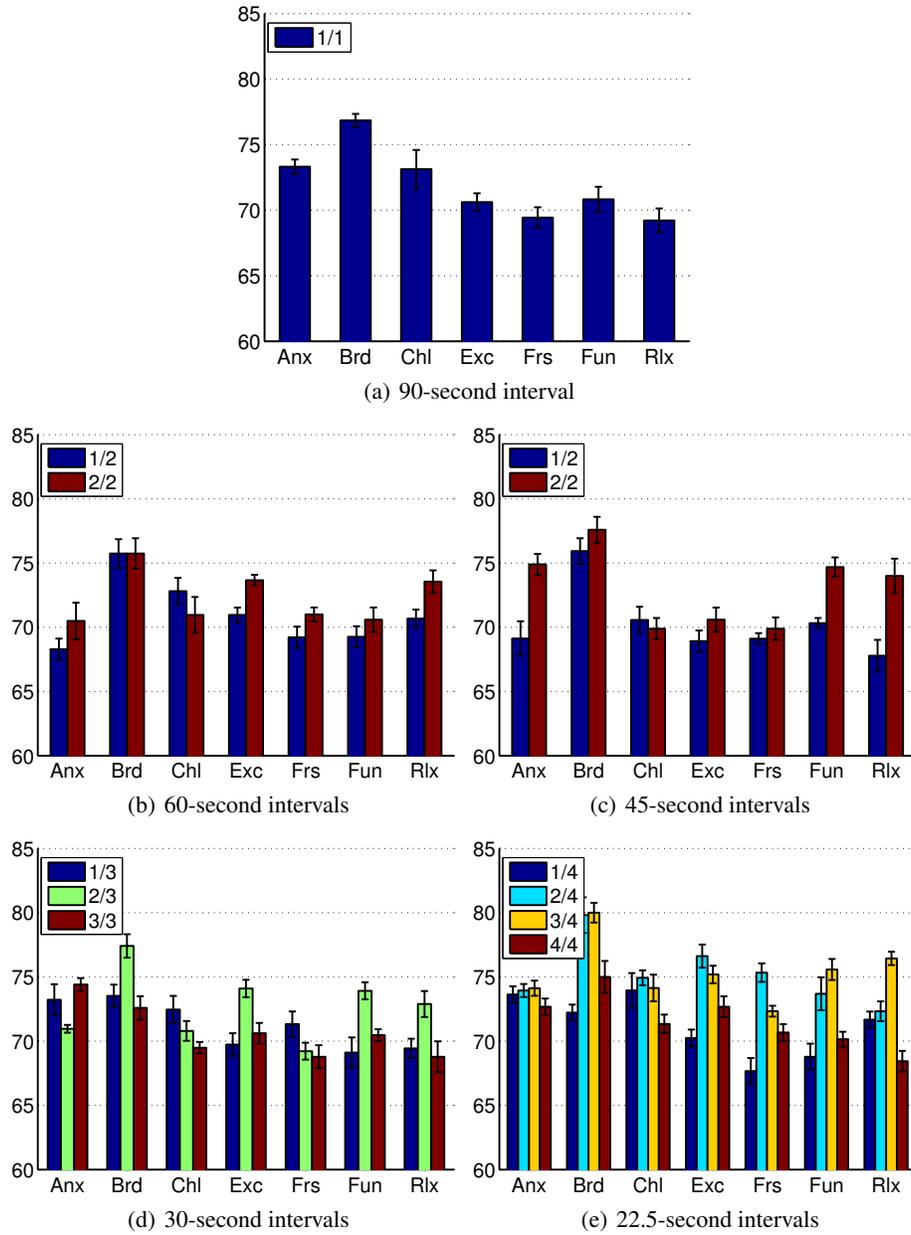
Fig. 1. Average and standard error of ten models trained on the 7 affective states — anxiety (Anx), boredom (Brd), challenge (Chl), excitement (Exc), frustration (Frs), fun (Fun) and relaxation (Rlx) — using statistical features calculated in different time-intervals of the experience.

accuracy in the last 60 seconds and in the first 30 seconds. This points out that self-reports of these two negative affective states are not significantly related to a specific part of the Maze-Ball experience and instead possibly linked to concrete asynchronous events in the game. It is possible that when players report fun, excitement, relaxation and boredom they build an overall assessment of the experience which is better approximated by the events occurring in the middle of the game while the reports of anxiety and frustration might be determined by punctual events in the game that elicited those affective states.

The predictors built for self-reports of challenge present, consistently across different time-window sizes, higher accuracies when trained on the first parts of the experience although no significantly different from the models trained on the complete 90 seconds of the game (higher difference equal to 1.81%; $t(18) = 1.18, ns$). It appears that in such short experience with a constant level of challenge — all elements in the level remain the same during each level — the first half of the game is enough to assess the difficulty of the level. Perceived challenge, opposite to the other six user states investigated, is not considered an affective state but rather a cognitive state which might be remembered more accurately than affect having an impact on which parts of the experience are taken into account while reporting the experience via questionnaires.

Despite the fact that only some states appear to be related to a specific interval of the experience, there exists at least one time interval for each of the seven states investigated that yields, on average, more accurate models than the complete game (not significant for challenge and anxiety). This suggests that statistical features calculated in small parts of the signals highlight better the differences on the experience than physiological features calculated in the full-length of the game.

6   Conclusions

The most common approach to create predictors of players' affective states based on their physiological states consists of applying machine learning algorithms to find mappings between reports of affect and statistical features extracted from the physiological signals recorded during play. This paper examines the effect of using different fractions of the physiological signals to the prediction of seven self-reported affective/cognitive states in a short 3D prey/predator game. Results show that features extracted from middle parts of the signals yield more accurate predictors of positive affective states and boredom than features extracted from the complete signals. Moreover, self-reports of perceived challenge are predicted more accurately by parts of the signals located at the beginning of the experience. Finally, reports of anxiety and frustration are not clearly related to a specific time window in the experience. Even though the games played are short (90 seconds), results show that more accurate models can be built when using a fraction instead of the complete physiological signals recorded during the game.

Future work will attempt to validate these results across more players and different casual games, i.e. short games with simple rules, that can provide comparable experiences. Additionally, this paper only uses information from the physiological state of the player leaving out game play events and game metrics which provide relevant information for the prediction of reported emotions (see [16] among others); an extended study

will explore whether the most informative intervals of that modality correspond to the same intervals of the physiological signals.

References

1. Gilleade, K., Dix, A., Allanson, J.: Affective videogames and modes of affective gaming: assist me, challenge me, emote me. In: Proc. DIGRA. vol. 2005 (2005)
2. Hudlicka, E.: Affective game engines: motivation and requirements. In: Proceedings of the 4th International Conference on Foundations of Digital Games. pp. 299–306. ACM (2009)
3. Mandryk, R., Inkpen, K., Calvert, T.: Using psychophysiological techniques to measure user experience with entertainment technologies. Behaviour & Information Technology 25(2), 141–158 (2006)
4. McQuiggan, S., Lee, S., Lester, J.: Early prediction of student frustration. Affective Computing and Intelligent Interaction pp. 698–709 (2007)
5. Yannakakis, G.N., Martínez, H.P., Jhala, A.: Towards affective camera control in games. User Modeling and User-Adapted Interaction 20, 313–340 (2010), 10.1007/s11257-010-9078-0
6. Picard, R., Vyzas, E., Healey, J.: Toward machine emotional intelligence: Analysis of affective physiological state. IEEE transactions on pattern analysis and machine intelligence pp. 1175–1191 (2001)
7. Cacioppo, J., Berntson, G., Larsen, J., Poehlmann, K., Ito, T., et al.: The psychophysiology of emotion. Handbook of emotions pp. 119–142 (1993)
8. Robinson, M., Clore, G.: Belief and feeling: Evidence for an accessibility model of emotional self-report. Psychological Bulletin 128(6), 934 (2002)
9. Kapoor, A., Burleson, W., Picard, R.: Automatic prediction of frustration. International Journal of Human-Computer Studies 65(8), 724–736 (2007)
10. Kivikangas, J., Ekman, I., Chanel, G., Järvelä, S., Salminen, M., Cowley, B., Henttonen, P., Ravaja, N.: Review on psychophysiological methods in game research. Proc. of 1st Nordic DiGRA
11. Nacke, L., Lindley, C.: Flow and immersion in first-person shooters: measuring the player's gameplay experience. In: Proceedings of the 2008 Conference on Future Play: Research, Play, Share. pp. 81–88. ACM (2008)
12. Rani, P., Sarkar, N., Liu, C.: Maintaining optimal challenge in computer games through real-time physiological feedback. In: Proceedings of the 11th International Conference on Human Computer Interaction. pp. 184–192 (2005)
13. Mandryk, R., Atkins, M.: A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. International Journal of Human-Computer Studies 65(4), 329–347 (2007)
14. Yannakakis, G.N., Hallam, J., Lund, H.H.: Entertainment capture through heart rate activity in physical interactive playgrounds. User Modeling and User-Adapted Interaction 18(1), 207–243 (2008)
15. Yannakakis, G.N., Hallam, J.: Entertainment modeling through physiology in physical play. International Journal of Human-Computer Studies 66(10), 741–755 (2008)
16. Martínez, H., Yannakakis, G.: Genetic search feature selection for affective modeling: a case study on reported preferences. In: Proceedings of the 3rd international workshop on Affective interaction in natural environments. pp. 15–20. ACM (2010)
17. Tognetti, S., Garbarino, M., Bonanno, A., Matteucci, M., Bonarini, A.: Enjoyment recognition from physiological data in a car racing game. In: Proceedings of the 3rd international workshop on Affective interaction in natural environments. pp. 3–8. ACM (2010)

18. Mcquiggan, S., Mott, B., Lester, J.: Modeling self-efficacy in intelligent tutoring systems: An inductive approach. User Modeling and User-Adapted Interaction 18(1), 81–123 (2008)

19. Broekens, J., Kosters, W., de Vries, T.: Eye movements disclose decisions in set. BNAIC (2009)

20. Conati, C., Maclaren, H.: Modeling user affect from causes and effects. User Modeling, Adaptation, and Personalization pp. 4–15 (2009)

21. Hazlett, R.: Measuring emotional valence during interactive experiences: boys at video game play. In: Proceedings of the SIGCHI conference on Human Factors in computing systems. pp. 1023–1026. ACM (2006)

22. Ravaja, N., Saari, T., Laarni, J., Kallinen, K., Salminen, M., Holopainen, J., Järvinen, A.: The psychophysiology of video gaming: Phasic emotional responses to game events. In: Proceedings of the DiGRA conference Changing views: worlds in play

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

24. Martinez, H.P., Jhala, A., Yannakakis, G.N.: Analyzing the Impact of Camera Viewpoint on Player Psychophysiology. In: Proceedings of the Int. Conf. on Affective Computing and Intelligent Interaction. pp. 394–399. IEEE, Amsterdam, The Netherlands (September 2009)

25. Tognetti, S., Alessandro, C., Bonarini, A., Matteucci, M.: Fundamental issues on the recognition of autonomic patterns produced by visual stimuli. In: Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on. pp. 1–6. IEEE (2009)

26. Fürnkranz, J., Hüllermeier, E.: Preference learning. Künstliche Intelligenz 19(1), 60–61 (2005)

27. Yannakakis, G.N.: Preference Learning for Affective Modeling. In: Proceedings of the Int. Conf. on Affective Computing and Intelligent Interaction (ACII09). Amsterdam, The Netherlands (September 2009)