Enjoyment Recognition From Physiological Data in a Car Racing Game

Simone Tognetti
Politecnico di Milano - IIT Unit
Dipartimento di Elettronica ed Informazione
Piazza Leonardo Da Vinci 32,20133 Milano, Italy
tognetti@elet.polimi.it

Maurizio Garbarino
Politecnico di Milano - IIT Unit
Dipartimento di Elettronica ed Informazione
Piazza Leonardo Da Vinci 32,20133 Milano, Italy
garbarino@elet.polimi.it

Andrea T. Bonanno
Simon Fraser University
School of Interactive Arts and Technology
250 -13450 102 Avenue
Surrey, BC V3T 0A3 Canada
abonanno@sfu.ca

Matteo Matteucci
Politecnico di Milano - IIT Unit
Dipartimento di Elettronica ed Informazione
Piazza Leonardo Da Vinci 32,20133 Milano, Italy
matteucci@elet.polimi.it

Andrea Bonarini
Politecnico di Milano - IIT Unit
Dipartimento di Elettronica ed Informazione
Piazza Leonardo Da Vinci 32,20133 Milano, Italy
bonarini@elet.polimi.it

ABSTRACT

In this paper we present a case study on The Open Racing Car Simulator (TORCS) video game with the aim of developing a classifier to recognize user enjoyment from physiological signals. Three classes of enjoyment, derived from pairwise comparison of different races, are considered for classification; impact of artifact reduction, normalization and feature selection is studied; results from a protocol involving 75 gamers are discussed. The best model, obtained by taking into account a subset of features derived from physiological signals (selected by a genetic algorithm), is able to correctly classify 3 levels of enjoyment with a correct classification rate of 57%.

Categories and Subject Descriptors

H.1.2 [**User/Machine System**]: [Human information processing]; I.5.2 [**Pattern recognition**]: Design Methodology—*Feature evaluation and selection, Classifier design and evaluation*

General Terms

Experimentation

Keywords

enjoyment classification, pairwise preference, torcs, physiological signals, artifact removals, data normalization, feature selection, SFS, genetic algorithms, k-nn ,cross validation

AFFINE'10, October 29, 2010, Firenze, Italy.

1. INTRODUCTION

Emotion has been investigated in the past by many researchers, including philosophers, psychologists, sociologists, psychophysiologists, and engineers. Results from psychophysiological studies [3, 16] indicate that relationships between the stimuli presented to a person and the observed physiological reactions may exists. Different types of stimuli have been used in the past such as movies [5] or images [12, 20]. Grounded on these findings, people working on Affective Computing aimed to design human-machine interfaces for real life application with emotion recognition abilities [8, 9, 18].

The emotion classification problem from physiological signals has been addressed in two main different ways: absolute and relative. The absolute approach aims to classify emotions experienced by people during a given stimulus and, in this case, the emotion associated to a stimulus is obtained from the averaging of self reported data [19, 5, 18]. The relative approach aims to predict the emotions from multiple comparisons of stimuli in a differential way. This second approach is named comparative affect analysis and it was first introduced by Yannakakis and Hallam [15, 14]; by means of preference learning [11, 13], each user reported preference is matched with changes in physiological signals between pairs of stimuli.

Regardless of the approach, to obtain a general, reliable and robust affect model, a number of fundamental issues should be considered. The protocol must ensure repeatability and minimize external interference; movement artifacts should be removed to improve accuracy; data have to be normalized to eliminate inter-subject variations (only in the case of first approach); features should be selected to reduce non-informative data and the obtained model has to be tested with a rigorous cross validation method to obtain a correct estimation of performance [16].

In this paper we present a method for enjoyment recognition from physiological data in a car racing game. To maximize performance and validity of the classification model we addressed a number of problems on acquired data such as inter-subject variability and movement artifacts. The enjoyment has been derived from the preference expressed by subjects between pairs of races. This keeps the advantages of the comparative affect analysis, but allows us to provide an absolute classification of the enjoyment. The video game scenario has been chosen since it allows high emotional involvement while keeping subject focused on the game. Performances of the classification method are reported for different preprocessed datasets. Impact of artifact reduction, normalization and feature selection is highlighted and results discussed.

In the next sections, we introduce the protocol designed for the experiments, and we present the methods we adopted for enjoyment modeling together with obtained results.

2. EXPERIMENTAL SETTING

We propose a new gaming experimental protocol, tested on a car-racing computer game, in which affective computing techniques can be applied and validated. The protocol was designed to produce an affective computing benchmark dataset, that could be used for developments and comparison. This dataset is composed by physiological data, questionnaire answers regarding user data (i.e., game experiences and race preferences), game logs and 2 video camera recordings. 75 volunteers (60 males and 15 females) aged from 18 to 30 (57 from 19-25, 18 from 26-30) took part to this study.

2.1 Task Design

The cognitive task in the experiment concerns playing a video game. This makes it possible to reach a high repeatability and a high level of involvement among participants. TORCS [21] was chosen as reference game for the following reasons: it is a video game that requires the player to be sitting in front of a computer, therefore subjects experiment emotionally different situations characterized by a similar physical activity and the effects of movement artifacts on acquired data are negligible, differently from what happened in [17] where the subjects had to move and jump; this game is an open source project, therefore, it has been possible to implement custom logging and AI for opponent drivers; it is easy enough, even for an inexperienced player, thus the game experience can be kept as homogeneous as possible among subjects involved in the experiment.

During a game session, each participant played 7 races versus one computer driver which is the only opponent during the race. The opponent skill has been changed among races considering that this has a high potential impact on player emotional state and that it could be easily adapted in a real-time affective loop to maintain the enjoyment level on the player according to the general principles of game engagement proposed by Malone [2]. Due to the previous observations, we assume that physiological responses are not task dependent since the level of physical activity required for the interaction with the game is constant during the session, while they do depends on the emotions felt during each race.

Three classes of game scenarios have been considered and a customized opponent driver has been implemented to match the skill of the player. It modulates its speed to keep a given distance from the human driven car. We call W (Winner) the driver that is more skilled than the player and that has the goal to keep a distance of +100 m (relative distance between the cars within the current track) from the player. C (Challenging) is the driver that is as skilled as the player and tries to keep a distance of 0 m from the player. Finally L (Loser) is the driver that is less skilled than the player and that keeps a distance of -100 m from the player. According to a priori considerations, the second variant of the race could be considered really challenging, and more interesting for the player. Race parameters such as type of track, environmental details, car model, and number of opponents have been chosen to keep the game easy to play and to make the opponent skill being the main difference among races.

2.2 Experimental Protocol

Most of the choices in the experimental protocol have been made to maximize the focus of the player on the task. The environment where the experiment took place has been conceived with the purpose of isolating the player and maximizing the game immersion so that no external event could influence the subject physiological state. The setting was a small room with a computer placed on a desk. Players were sitting in front of the monitor and they were interacting with the computer through standard mouse and keyboard. No other people were in the same room and the operator monitored the experiment from an external site.

Before the experiment, all participants have been asked to fill out individually a general questionnaire, presented in computer-based form, used to gather information about their experience with video games, game preferences, TORCS prior knowledge, and personal data such as age and handedness.

Participants have been fitted with sensors to measure peripheral physiological activity as explained in Section 2.3. The players were asked to wear a headphone to guarantee a deeper game involvement through race sounds. After this setup phase, the player was left alone listening to a relaxing music (i.e., sounds from nature) with the purpose of decreasing the stress and the initial excitement for the test and of bringing all the subjects to a similar starting condition. Cameras and physiological signal acquisition were started while the player was waiting. The subjects have been instructed to minimize movements during the task to avoid artifacts.

To increase subject involvement during the game, players have been told that they were competing for a prize. Prizes were given basing on a series of parameters including in-game performance, but also on physiological features, so that potential advantages of skilled player were reduced. Note that, from this moment on, to avoid the effect of covert communication [10], no further interaction between operator and subject occurred. The protocol was carried on by an automatic script on the computer that started each race and managed the questionnaire (see Section 3).

After about a minute of relaxing music, the participants were asked to read the instructions and then, to start the experiment by pressing a button. At the end of each race, starting from the second one, the participants were asked by a script to express, via a computer-based form, the preference between the race just played and the previous one. To minimize any potential order effect on physiological and self-reported data, each pair of game variants have been presented in both orders. The sequence of driver classes was as follows: W C L W L C W. The duration of each race was 3 minutes; this provided enough time to eliminate past race effects on physiological signals and to produce a new arousal level before the overcoming of boredom caused by excessive race length. The total time of a session was about 30 minutes, i.e., 21 minutes (7 races × 3 min.) of racing and about 7 minutes of setup, question answering and resting.

2.3 Acquired data

In this protocol four types of data have been acquired: physiological data, questionnaire answers (presented in Section 3), game logs and video camera recordings. All different sources were synchronized to ensure a correct multimodal analysis.

Physiological data were gathered using the ProComp Infiniti device. This device captured 5 physiological signals: Blood Volume Pulse (BVP), Electrocardiogram (ECG), Galvanic Skin Response (GSR), Respiration (RESP) and Temperature (TEMP). A sample rate of 256 Hz has been used except for ECG and BVP signals that were sampled at 2048 Hz. The hand not used for interacting with the game was fitted with GSR, BVP and TEMP sensors. The 3 terminal ECG sensor were placed on the chest while the RESP sensor was placed around the chest.

Based on previous literature [19, 20, 8, 17], several derived signals have been extracted from the basic ones at the same sampling rate. Heart Rate (HR) has been derived both from ECG (HR_{ecg}) and BVP (HR_{bvp}); magnitude (SM) and duration (SD) of signal variation has been derived from GSR; inspiration/espiration time ($inTime$, $outTime$), apnea in/out time ($apneaup$, $apnealow$) and respiration interval ($rTime$) have been extracted from respiration signal; upper/lower envelope of BVP (BVP_{up},BVP_l) and their difference ($BVP_d = BPV_{up} - BVP_l$) have been also computed.

The feature vector $F = [f_1 f_2 \dots f_D] \in \mathcal{R}^D$ has then been obtained by the union of a set of features computed for each of the mentioned signal during each race. The used features are: mean, variance, min and max value and their difference, time of min and max value, delta time between min and max value, first and second differences, trend and auto correlation function. So we have 16 signals (basic+derived except the raw ECG) and 12 features that gives $D = 192$. We assume that the first part of each race is subject to transitory phenomena due to the transition from a race to the next one. Thus, the features have been computed by considering only the last 60 seconds of each race.

Two video cameras recorded the environment in which the player acted too; but, these data have not been considered in the analysis presented in this paper, and will be used by further research activities.

3. QUESTIONNAIRE ANALYSIS

Preferences between races have been collected. At the end of each race, the subjects were asked whether they enjoyed more the last race or the previous one. A pairwise preference scheme (2-alternative forced choice: 2-AFC) has been used in self reports. 2-AFC offers the main advantage of acquiring objective enjoyment: it normalizes the different perception of enjoyment among subjects and it allows a fair comparison between the answers of different subjects. Being interested in a general model linking physiological features and reported entertainment preferences (which generalizes over different players), an approach robust with respect to users, like 2-AFC, has to be preferred instead of approaches involving direct ranking [1].

In Table 1, the distribution of users answers is reported. Each answer is represented by a set of bits where 0 or 1 indicates whether the subject preferred previous or current race respectively. The first bit indicates the preference between the first and second race. For each preference sequence we indicate, the total number (and frequency) of people answered in that way.

Table 1: Questionnaire answer distribution

Answer	Count	Freq.	Answer	Count	Freq.
000000	1	0.013	100110	6	0.080
001001	1	0.013	100111	1	0.013
001101	1	0.013	101010	32	0.427
010111	1	0.013	101011	3	0.040
011001	1	0.013	101110	9	0.120
011010	1	0.013	101111	1	0.013
100010	5	0.067	110110	3	0.040
100011	1	0.013	110111	1	0.013
100100	2	0.027	111010	2	0.027
100101	2	0.027	111100	1	0.013

To derive the classes used as ground truth in the enjoyment recognition task, we need to derive a discrete value of enjoyment (e.g., high, medium or low) to each race from the result of Table 1. This has been obtained by implementing the following voting mechanism: for each possible race permutation (i.e., possible ranking), all pairs of races for all subjects are considered and a positive vote is given for each comparison that results to be coherent with respect to the current ranking.

For instance, let's consider the ranking (from best to worst) $R = [3, 2, 4, 7, 1, 6, 5]$ and the questionnaire answers $A = [101010]$. The first user answer (i.e., A[0]=1) means that the player preferred the current race with respect to the previous, i.e., race 2 is preferred with respect to race 1. Since for the rank R we are considering that 2 is better than 1, a vote is given to R. The second answer (i.e., A[1]=0) means that the player preferred the previous race to the current thus, race 2 is preferred with respect to race 3; however being 3, in the ranking R, better than 2, no vote is given for the second answer. The analysis of data in Table 1 lead to the following ranking {6,2}, {4}, {7,1}, {3,5} where races 6 and 2 are the favorite while 3 and 5 are the least favorite (brackets group races of equal rank). It is possible to group races to obtain the following 3 discrete level of enjoyment: {6,2} High, {4,7,1} Medium and {3,5} Low.

4. SIGNAL PREPROCESSING

The signal preprocessing aims to improve the accuracy of the enjoyment classifier through two main signal elaboration techniques: artifact removal and data normalization. Artifact removal aims to remove artifact due to sensor movements that might lead to a misclassification. Data normalization aims to remove variability among subjects data (i.e., feature have different means among subjects) to improve classification performance with different subjects.

4.1 Artifact Removal

One of the main problem affecting the quality of physiological signals used in this experiment are artifacts due to movements during game play: even if the interaction involves only the hand with no sensors, often the player moves

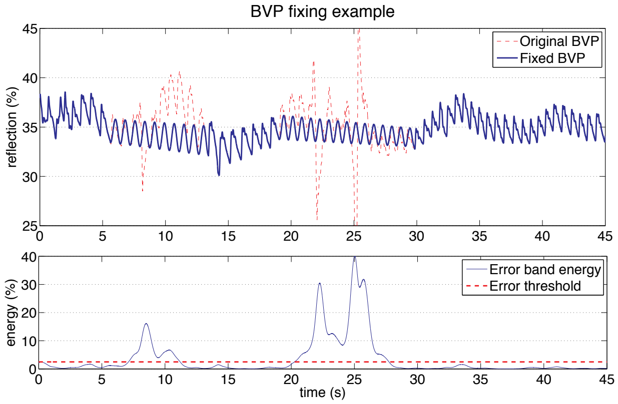

Figure 1: An example of BVP fixing. The artifact is replaced with an appropriate sinusoidal function.

the other arm causing errors in acquisition and noise in the data. Resulting artifacts need to be removed.

Signals mostly affected by this problem are BVP and ECG, being their respective sensors very susceptible to this kind of disturbance. The characteristic shape and trend of both ECG and BVP has to be preserved, since its poor quality could affect the analysis of signals which derive from it. In particular, to extract HR, the typical shape of an heart beat signal (QRS complex) in ECG has to be maintained as well as BVP frequency and variations that are evaluated to measure systolic and diastolic pressure and HR.

Our approach analyzes the signal Power Spectral Density (PSD) with the purpose of identify intervals where the signal is affected by noise. The signal spectrogram is first computed to estimate signal energy, within a fixed band, at different times. The energy is then compared against a threshold. Each time steps of the signal is then labeled as noisy if the estimated energy, in the selected band, is higher than the given threshold. The selected frequency band is also called error band since it is the most affected by movement artifacts. The threshold and the frequency of the error band have been manually chosen after an intensive analysis of noisy signals by looking at the frequencies in which signal was most affected by noise.

After noisy parts of signals have been detected, a fixing algorithm (or reconstruction algorithm) is used to replace the corrupted signal with a new one that have no errors. While the noise detection method is similar between different signals (although error frequencies might change), the reconstruction process is different accordingly to the signal that have to be fixed.

For each noisy segment, the BVP fix algorithm starts by computing BVP's intrinsic parameters like the mean heart rate and the mean values of upper and lower levels of the signal before and after the noisy part. Then, the corrupted portion of the signal is replaced by a sine function defined by these parameters: the amplitude is derive by values of upper and lower levels and the period depends by the mean heart rate. The result of a fixing operation on BVP is reported in Figure 1.

In the case of ECG fixing, the mean heart rate is computed in non noisy segments adjacent to the corrupted one. Then, the characteristic waves of ECG are reproduced, filling the noisy segment, with fixed distance depending on the measured heart rate.

For what concerns other signals, artifacts are simply removed by applying suitable filters. This is the case of the respiration signal which may have different mean values according to the posture assumed by subject. Since this might produce a misclassification of data we apply a 0.04Hz high-pass filter to this signal.

Finally, GSR and Temperature might be affected by high frequencies noise that can be easily removed with a 0.4Hz low pass frequency filter.

4.2 Normalization

Physiological signals are characterized by large inter-subject variability that should be reduced when we are looking for a general model of affect that holds for many people (e.g., in general, the resting HR of an athlete is slower compared to the one of an out of training person). Data are affected by inter-subject variability when they are distributed over different subspaces of the D-dimensional feature space R^D. In this situation, a data normalization technique is required to transform subjects distributions into a standard space in which classification can be performed. If normalization is not applied, the classifier could not be able to generalize over new data resulting in a low classification accuracy.

Normalization should be used after artifact removal, otherwise the results of this process might depend on the extent of errors affecting the signals. Moreover, normalization should not reduce the information related to the classification problem, thus it is applied independently to each signal of each subject, merging together all the data.

By means of kernel based methods the probability density of each signal is estimated. This is fundamental since it allows a further outliers detection of data that lie on the probability density tails. Values with an estimated density lower that a given threshold are cut off. The normalization process scales each signal to a $[0, 1]$ interval through $\bar{s}(t) = \frac{s(t)-s(t)_{min}}{s(t)_{max}-s(t)_{min}}$ where $s(t)$ and $\bar{s}(t)$ are the original and normalized signals, while $s(t)_{min}$ and $s(t)_{max}$ are the upper and lower cut off boundaries obtained by the density estimation.

5. DATA ANALYSIS

In this section, the performance of the model for predicting enjoyment is discussed. We decided to use a K-NN (K=1) classifier, since it is one of the most robust and it converges to the Bayes optimal classifier. In order to have a correct estimation of performances, we applied a leave-one-subject out cross validation schema. This method uses the data of one subject for testing and data from remaining subjects training. Correct classification rates (CCR) are computed for each test subject and averaged.

Two types of analysis have been carried out. The first one, aims to evaluate the improvement that could be obtained in terms of CCR through the artifact removal and normalization while classifying enjoyment from a single feature. The second one aims to establish the improvement in terms of CCR produced by adding feature selection.

5.1 Impact of data normalization and artifact removal

Figure 5.1 shows the comparison among accuracy values

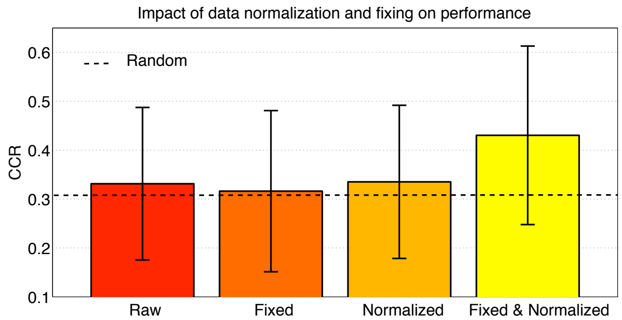

Figure 2: Impact on performance of fixing (artifact removal) and normalization of data applied to the best performing feature ($CCR = 43.05\%$). Anova test indicated that there is a significative difference between the mean of normalized and fixed feature with respect to others preprocessing techniques

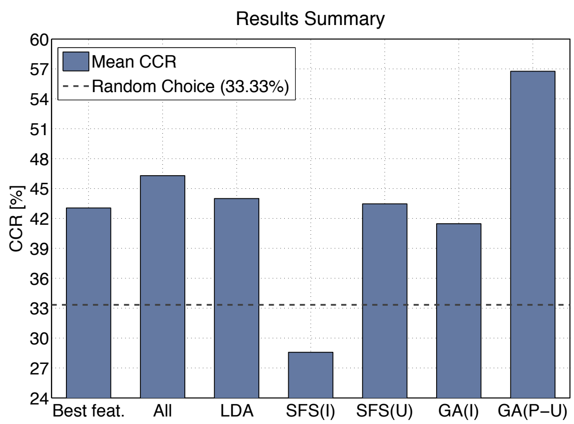

Figure 3: Comparison of results obtained by different feature selection algorithm

achieved on data from differently preprocessed data sets: raw data, fixed data, normalized data and finally, fixed and normalized data. The figure shows values of mean CCR achieved by the best performing feature on the normalized and fixed dataset: bvp_v. The artifact removal and normalization alone do not improve the performance of row data. However, combination of outlier removal and normalization increases accuracy to 43.05%. We have also performed an anova test that indicated a significative difference between the mean of fixed and normalized dataset w.r.t the other types of preprocessing techniques. Moreover, performance of bvp_v were initially close to the one of a random classifier but, after preprocessing, bvp_b became the best performing feature. We can conclude from this analysis that artifact removal, together with normalization, are fundamental for preserving and enhancing the information in the data.

5.2 Impact of feature selection

In this section we used the normalized and de-noised data set in order to test whether performance can be further improved by feature selection or projection. We started by considering the performance of a classifier built over all the features. Results indicated that a K-NN classifier trained and tested on complete data achieves a performance value with $CCR = 46.43\%$ that is slightly higher from the single feature performance.

We then applied Linear Discriminant Analysis (LDA) to find a linear transformation of features able to retains the maximum amount of class discriminant information in a lower-dimensional space. Therefore, we expect that the mean accuracy might increase. Once again the leave-one-subject-out cross-validation has been applied: LDA is trained on the training data set and the performance of the model is tested on the data left out from training. Using LDA, the classifier performance decreases to $CCR = 44\%$. LDA can be applied when data are linearly separable; thus, this performance reduction may suggest that data are not linearly separable.

To find the minimal subset of features that maximize the classification performance, we have evaluated and compared a greedy approach and an optimal technique for feature selection [7]: Sequential Forward Search (SFS) [6] and Genetic Algorithms (GA) [4]. SFS is a fast bottom-up algorithm that, at each step, adds an additional feature to the current feature set. The selected feature is added if its marginal value with respect to the classification function is positive. The GA has an higher computational demand and manipulates a population of individuals in binary encoding: each individual represents a set of features whose success in life is a function of its discriminative power. When GA is used for feature selection, the fitness function is the mean CCR obtained over selected features.

To evaluate the performance of feature selection algorithms, we have used an external cross-validation method. To avoid selection overfitting a, k-fold (k=15) cross-validation by subject has been performed. Data belonging to all the players in the original datset D_s have been randomly split into k folds D_{s_i} with i=1...15 (containing data of 5 subjects each). Folds have been assembled into 15 data set containing training data $D_{s_i}^{tr}{=}D_s \setminus D_{s_i}$ and validation $D_{s_i}^{val}{=}D_{s_i}$. The feature selection algorithm has been executed k times independently over each different data set. For each iteration, the best feature selection with respect to the training data $D_{s_i}^{tr}$ has been evaluated on testing data $D_{s_i}^{val}$. Each selection found on training has been evaluated with the same leave-one-out cross-validation procedure described previously.

Finally, features selected by the presented feature selection techniques have been combined in different ways: Pseudo-Union, Union and Intersection. Through a Pseudo-Union, a feature is selected if it is present in more than 50% of solutions, that for GA means at least 8 chromosomes extracted. This approach can tend to overfit data from the single set as long as the threshold used to compose final solution is less selective. The union is a Pseudo-Union with low threshold so that all features that are selected at least once are considered. By performing an intersection, a feature is selected only if it is in all different solutions. This strategy is the most correct, since the solution considers features that are found relevant for all sets. In Figure 5.2 are summarized the results of the data analysis. The intersection of features selected by SFS (mean of SM and SD) poorly performed with a $CCR = 27.43\%$, even lower than random chance. The union of all features selected by SFS (21 features of which most relevant are: variance of BVP, mean of HR, variance of GSR, mean of $outTime$), with $CCR = 43.32\%$, presented slightly better performances to the intersection from the GA, $CCR = 41.79\%$ (5 feature from BVP, RESP and *apnealow*). The best performances of $CCR = 56.95\%$ were achieved by the Pseudo-Union of features selected by the GA (91 features from almost all signals).

6. CONCLUSIONS

In this paper we presented a methodology for the enjoyment classification from physiological signals in car racing game and games in general.

A few fundamental aspects of emotion classification have been discussed and addressed: the protocol has been designed to properly stimulate the subjects in a way that different emotional states could be elicited; a 2-AFC questionnaire has been used to acquire ground truth data from differential preference reported by users; physiological signals have been processed and, from resulting features, a classifier has been derived.

The methodology used to acquire ground truth preference data, i.e., the preference ranking, has shown an advantage in accuracy respect to a direct ranking questionnaire being able to normalize responses with respect to user subjective enjoyment perception. In addition, the ranking method applied on the questionnaire has allowed to determine the presence of three enjoyment classes for a classification problem.

Two general techniques for data preprocessing have been discussed: physiological signals have been processed to remove any artifact due to user movement or noise and a normalization process has been applied with the purpose of reducing inter-subject variability.

To ensure a general validity of the K-NN classifier, reported results have been validated with a cross-validation method. A first data analysis has been used to measure the accuracy achieved with a single feature classifier. This result has been further improved by means of feature selection algorithms. In the future, we might consider trying other classifiers such as neural network, support vector machines or decision trees.

Given the complexity of the task, 56.95% of accuracy over 3 classes is a considerable result. However, by analyzing the confusion matrix, we observe that there is a consistent overlap of classes W and L that partially explains such performance. This is also confirmed by subject's self reports that indicate only a strong preference of class C over the others.

7. ACKNOWLEDGMENTS

The research activity described in this paper has been partially supported by IIT - Italian Institute of Technology.

8. REFERENCES

[1] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 140:1–55, 1932.

[2] T. Malone. What makes computer games fun? In *Proceedings of the joint conference on Easier and more productive use of computer systems.(Part-II): Human interface and the user interface-Volume 1981*. ACM New York, NY, USA, 1981.

[3] P. Ekman, R. Levenson, and W. Friesen. Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616):1208–1210, 1983.

[4] D. Goldberg and J. Holland. Genetic algorithms and machine learning. *Machine Learning*, 3(2):95–99, 1988.

[5] J. Gross and R. Levenson. Emotion elicitation using films. *Cognition & Emotion*, 9(1):87–108, 1995.

[6] D. Aha and R. Bankert. A comparative evaluation of sequential feature selection algorithms. *Springer-Verlag, New York*, 1996.

[7] M. Dash and H. Liu. Feature selection for classification. *Intelligent data analysis*, 1(3):131–156, 1997.

[8] R. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, pages 1175–1191, 2001.

[9] C. Lisetti, F. Nasoz, C. LeRouge, O. Ozyer, and K. Alvarez. Developing multimodal intelligent affective interfaces for tele-home health care. *International Journal of Human-Computer Studies*, 59(1-2):245–255, 2003.

[10] R. Rosenthal. Covert communication in laboratories, classrooms, and the truly real world. *Current Directions in Psychological Science*, pages 151–154, 2003.

[11] J. Furnkranz and E. Hullermeier. Pairwise preference learning and ranking. *Lecture Notes in Computer Science*, 2837:145–156, 2003.

[12] I. Christie and B. Friedman. Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology*, 51(2):143–153, 2004.

[13] J. Doyle. Prospects for preferences. *Computational Intelligence*, 20(2):111–136, 2004.

[14] G. Yannakakis, H. Lund, and J. Hallam. Modeling children's entertainment in the playware playground. In *Proc. of CIG*, pages 134–141, 2006.

[15] G. Yannakakis and J. Hallam. Towards capturing and enhancing entertainment in computer games. *Lecture Notes in Computer Science*, 3955:432, 2006.

[16] J. Cacioppo, L. Tassinary, and G. Berntson. *Handbook of Psychophysiology, 3rd ed.*. Cambridge University Press, New York, NY, 2007.

[17] G. Yannakakis and J. Hallam. Entertainment modeling through physiology in physical play. *International Journal of Human-Computer Studies*, 66(10):741–755, 2008.

[18] J. Kim and E. André. Emotion recognition based on physiological changes in listening music. *IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 30 (12), pp. 2067-2083, December*, 30(12):2067–2083, December 2008.

[19] A. Bonarini, L. Mainardi, M. Matteucci, S. Tognetti, and R. Colombo. Stress recognition in a robotic rehabilitation task. In *Proc. of "Robotic Helpers: User Interaction, Interfaces and Companions in Assistive and Therapy Robotics", a Workshop at ACM/IEEE HRI 2008*, volume 1, pages 41–48, Amsterdam, 2008. University of Hertfordshire.

[20] S. Tognetti, C. Alessandro, A. Bonarini, and M. Matteucci. Fundamental issues on the recognition of autonomic patterns produced by visual stimuli. In *Proc. of ACII 2009*, Amsterdam, Netherland, 2009.

[21] The open racing car simulator website. http://torcs.sourceforge.net/.