

A Psychophysiological Evaluation of an Effective Recognition Technique Using Interactive Dynamic Virtual Environments

Mohammadhossein Moghimi, Robert Stone, Pia Rotshtein

Abstract—Recording psychological and physiological correlates of human performance within virtual environments and interpreting their impacts on human engagement, ‘immersion’ and related emotional or ‘effective’ states is both academically and technologically challenging. By exposing participants to an effective, real-time (game-like) virtual environment, designed and evaluated in an earlier study, a psychophysiological database containing the EEG, GSR and Heart Rate of 30 male and female gamers, exposed to 10 games, was constructed. Some 174 features were subsequently identified and extracted from a number of windows, with 28 different timing lengths (e.g. 2, 3, 5, etc. seconds). After reducing the number of features to 30, using a feature selection technique, K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) methods were subsequently employed for the classification process. The classifiers categorised the psychophysiological database into four effective clusters (defined based on a 3-dimensional space – valence, arousal and dominance) and eight emotion labels (relaxed, content, happy, excited, angry, afraid, sad, and bored). The KNN and SVM classifiers achieved average cross-validation accuracies of 97.01% ($\pm 1.3\%$) and 92.84% ($\pm 3.67\%$), respectively. However, no significant differences were found in the classification process based on effective clusters or emotion labels.

Keywords—Virtual Reality, effective computing, effective VR, emotion-based effective physiological database

I. INTRODUCTION

THE recent “resurrection” of interest in Virtual Reality (VR), courtesy of new interface and gaming technologies, many evolving from international crowd-funding communities has, once again, stimulated interest in the quest for true “immersion” or the generation of a believable sense of “presence” in computer-generated worlds. Although, Human-Computer Interaction (HCI) system designers have, in their attempts to increase the sense of end user immersion, introduced several multi-dimensional input/output devices, in an order to provide user-friendly, intuitive techniques and styles of interaction with real-time 3D worlds (including various types of data input controllers, multifunctional touch panels, for example); the area of HCI research that strives towards establishing direct communication between a computer system and the human brain has, until recently, been treated as science fiction (referencing such popular films as

The Matrix and Pacific Rim). In 2006, Cairns suggested that true “immersion” may only ever be achieved through the use of advanced brain-computer interfaces [1]. However, until that day arrives, it is important to understand in advance, how it may be possible to measure and, indeed, influence human engagement and emotional connectivity with virtual worlds using psychophysiological techniques.

In the VR domain, Brain-Computer Interaction (BCI) systems attempt to improve human-computer interaction and increase the sense of immersion by interfacing directly with the human brain and, thus, removing the artificial barriers to intuitive interaction afforded by conventional input-display techniques. So far, the interaction process has been mostly based on conventional methods, in that computer users typically use physical interaction devices to see, hear, act, sense haptic or olfactory stimuli, and in some cases, even talk to the system. The near-term goal of BCI systems, as an extension to these conventional systems (as opposed to a replacement, which is a longer-term aspiration), would be to translate human thoughts and emotions by direct connection to the human brain and use this information as a new modality channel for HCI systems [2]

As discussed in a previous paper by the present authors [3], to date, researchers have studied the implementation of virtual realities in many different areas. As well as entertainment, virtual realities and their so-called “serious games” counterparts have been used for training purposes [4]-[6], pain distraction [7], [8], rehabilitation régimes [9], [10] and emotional disorder therapy [11], [12]. The focus of all these studies has been to engage the human users in an interactive virtual environment, and to increase the sense of presence and immersion within them, thereby effectively delivering new skills, knowledge or in some cases, acting as a form of clinical distraction. In 2006, Joels suggested that changes in the excitement level (depending on the pleasurable or displeasurable condition), affects the learning and memory process. He proposed that memory performance changes (either improvements or impairments) are highly dependent on the time and context of the emotional experience [13]. Therefore, the recognition of the users’ emotions, when exposed to virtual realities, and controlling their affective experiences within the virtual environments (regardless of their purpose) can be as important as the VR’s contextual outcome.

As highlighted by the present authors in [3], one of the sub-categories of research into BCI systems is described as

Mohammadhossein Moghimi and Prof. Robert Stone are with the Department of Electronic, Electrical and Systems Engineering, University of Birmingham (e-mail: m. moghimi@pgr.bham.ac.uk, r.j.stone@ bham. ac.uk).

Dr. Pia Rotshtein is with the School of Psychology, University of Birmingham (e-mail: p.rotshtein@bham.ac.uk).

affective computing. During the process of affective computing, psychophysiological signals from the users are recorded to enable the BCI system to extract data of relevance to their emotional and cognitive states. This new input channel could provide several features for an advanced HCI system attempting to support the generation of believable immersive experiences. As an illustration, the system could use this information to adapt itself to the user's emotions and, by doing so increase his/her performance and immersion levels, during the interaction process. Recently, developments in HCI-mediated emotional recognition have been developed using non-interactive, or passive environments, such as listening to music, or the observation of videos and imagery (e.g. [14]-[23]), with others beginning to focus on virtual realities and more interactive environments (e.g. [24], [25]).

In [3], we conceptualised, designed and evaluated an Affective Virtual Reality (Affective VR), capable of evoking various emotional experiences on the part of the human user. In the present study, by employing the designed Affective VR, an affective computing system was designed and evaluated. To do this, the relationship between psychophysiological signals and human emotions, evoked through the designed Affective VR (presented in [3]), has been the focus of investigation.

II. RELATED WORKS

A. Affective Stimuli and Experience Assessment

As described in a previous paper by the present authors [3], to analyse the emotional response of humans and their psychophysiological responses, a psychophysiological affective database, recorded from a number of users exposed to a number of controlled and known affective stimuli, is required. To construct such a database, a number of controlled emotional scenarios (affective stimuli) evoking some specific affective states on the part of the users need to be presented to participants in an experiment, whilst taking part in a physiological measurement paradigm. These recordings, tagged by the corresponding affective states, are analysed for the design, training and validation of the affective recognition system. So far videos [14], [15], music videos [16], [17], Images [18], [19], Sound [20], [21], Real Life Scenarios [22], [23] and Virtual Reality and Games [24], [25] have been employed, as affective stimuli, in order to evoke a range of emotional experiences, on the part of the users. On the other hand, studies have employed either self or expert assessments to tag the emotional stimuli with the participants' affective experiences. In expert-assessment a psychologist or human emotion expert is instructed to evaluate the participant's affective state, and to categorise it within an Affective Space [23]. Whereas in self-assessment, the participants were instructed to evaluate their emotional experience and report them within an Affective Space [14]-[19], [21], [25]. To date, studies, in the main, have employed either dimensional (used by [14]-[16], [18], [19], [21], [24]) or categorical (employed by [23], [25]) Affective Space, to perform emotional experience assessments. In dimensional models, a number of

parameters are employed to numerically present emotional experiences within a dimensional space. Both Russell and Mehrabian presented two similar dimensional models in the 1980s and 1970s. These models define emotions based on two or three continuous independent parameters (Valence, Arousal and Dominance) [26], [27]. Whereas in categorical models, the Affective Space is presented by using an emotion set (a number of 'Emotion Labels'), such that the user can be "categorised" as experiencing either one or a combination of these Emotion Labels. As an illustration, Ekman and Friesen used a categorical presentation of emotions, labelling them as surprise, fear, disgust, anger, happiness and sadness [28].

B. Physiological Recordings and Features

To record psychophysiological responses of the users, exposed to affective stimuli (image, video, etc.), various physiological recordings have been employed in the literatures. To date, Electroencephalography (EEG) [14]-[18], [20], [24], [25], [29], Galvanic Skin Response (GSR) [16]-[19], [22]-[24], [29] and Heart Rate [16], [17], [19]-[24] have been the most popular recordings in the literature suggested to be related to affective states. However, a minority of the studies has also employed respiratory (breathing) rate, skin temperature, Electromyography (EMG) and pupil diameter, as well, in order to classify affective states [14], [16], [17], [24].

To train the emotion recognition agent (considering supervised learning algorithms [30]), a number of psychophysiological features need be identified and extracted from the recorded physiological signals. These features need to be related to the affective states, as they will ultimately be employed within the affective recognition system to predict the emotional response of the users when experiencing a specific affective situation. To date various physiological features have been introduced in the literatures. These features can be:

1. The statistical analysis (e.g. mean, standard deviation, etc.) of the raw signals (e.g. average GSR value, mean of the heart rate peaks, etc. [24]).
2. The frequency analysis of physiological signals to extract specific rhythms (e.g. alpha, beta, gamma rhythm powers within EEG signals [14], [17]).
3. The detection of specific patterns, such as Event Related Potentials (ERP – such as the P300, N100, and others [18]).
4. Other exclusive measurements (e.g. EEGw [29]).

III. PSYCHOPHYSIOLOGICAL DATABASE CONSTRUCTION

A. Material

In the present study to construct the psychophysiological database, the designed and pre-evaluated Affective VR, presented in [3], has been used as the source of the emotional stimuli. The Affective VR was based on a speedboat simulation (Fig. 1) acting as the background scenario. A number of parameters (called affective incidents) were implemented in the VR to change the affective power of the environment within the Circumplex of Affect presented by

Russell in 1980s [26]. As an illustration, participants were challenged by driving the boat and collecting scores freely, in a minefield or whilst being targeted by torpedoes, in various experimental setups, such as coloured images, black and white or inverse black and white screens, using a mouse, or a joystick, with or without simple force feedback (for more details, refer to [3]).

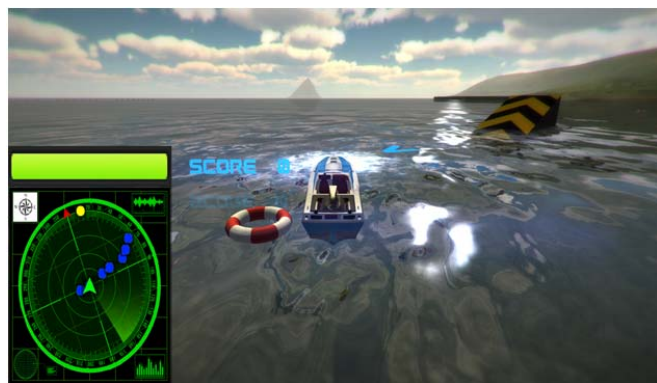


Fig. 1 Speedboat Simulation Environment

In the present study, the two most powerful affective games, in each of the four Affective Clusters introduced in [3], have been identified using the Cosine Similarity Algorithm [31] as implemented in [3]. As a result of this analysis, the eight most powerful affective games (those, which have the highest probability of driving the emotional experience of the participants toward all affective clusters) have been identified. Following the identification of the most powerful affective games, two neutral games were added in the experiment (the neutral game from [3], plus the game close to (0, 0, 0) with the highest standard deviation). Therefore, overall, 10 affective games have been identified for presentation to the participants in this experiment.

As discussed in [3], Multi-Variant Analyses of Variance (MANOVA) highlighted significant differences between the four participant groups (male gamers, male non-gamers, female gamers and female non-gamers). According to the results presented in [3], male gamers, male non-gamers and female gamers show marked similarities in their affective experiences, when compared to female non-gamers. In order to minimise between participants variability, it was decided to recruit only **male and female gamers** in this experiment

As the majority of studies have employed EEG, Heart Rate and GSR signals to perform affective analysis and recognition (Section II A), in the present study, it was decided to record data using these three techniques, for the purposes of supporting the psychophysiological database construction process. Participants were required to wear an EMOTIVE EPOC (developed by EMOTIV Inc.) headset to record EEG signals, as well as Shimmer+ wearable sensor technologies (developed by Shimmer Sensing Inc.) to record GSR and heart rate activities. The EMOTIVE EPOC records the EEG signals, with a 128Hz sampling frequency, from 14 channels (AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1 and O2), while P3 (Common Mode Sense – CMS) and P4 (Driven

Right Leg – DRL) are used as the reference channels), arranged according to the 10-20 EEG system. The GSR and heart rate data are also recorded using the Shimmer+ wearable sensor technologies, with a 512 Hz sampling frequency. A program was developed to function in parallel with the game, to perform the recording (through the software development tool kits (SDKs) provided by the manufacturers), as soon as each game was started.

B. Method

The experiment was performed in a quiet room. All participants were provided with a 32-inch Samsung HD LCD display, a Microsoft Wireless Mouse 5000, a Logitech Wingman 3D force feedback joystick and Sennheiser earphones. Each experiment commenced with a training session to prepare the participants for every possible incident within the games (as presented in [3]). The training introduced the game environment to the participants and served to reduce any element of surprise in the games. After the participants had completed the training session, they progressed to the two neutral games, followed by the other eight in a random order. At the end of each game the participants were instructed to self-assess their average emotional experience, based on the dimensional (Valence, Arousal and Dominance) and the categorical (according to eight Emotion Labels: Relaxed, Content, Happy, Excited, Angry, Afraid, Sad and Bored) models of affect (as presented in [3]). The participants were given a five-15 minute break, after playing the first five games, in order to reduce the fatigue factor caused by wearing the physiological sensing equipment. On average, each game lasted for three minutes, and the complete experiment took approximately 1.5 hours.

C. Result (Psychophysiological Database)

Of a possible total of 300 affective sessions, 290 were recorded (10 sessions were not attended by participants). During the affective sessions, the raw EEG signals from all 14 channels were recorded. Furthermore, the signal quality of each EEG channel was available from the EMOTIVE EPOC headset and was therefore recorded alongside the raw channel data. The raw Photoplethysmogram (PPG) output was recorded by the Shimmer+ device, mounted on the participant's index finger. During this recording a location of the skin is illuminated, and then the changes in light reflection are recorded. The alternating current component of the PPG signal relates to the blood pulse pressure. The Shimmer+ software uses the estimation techniques introduced in [32] to approximate the heart rate of the subjects using the PPG signal. Moreover, the GSR signal was also recorded using two finger straps mounted on the middle and ring fingers. These raw data sources were synchronised according to the master clock of the main system and stored in Microsoft Excel files during the run-time of the experiment. The emotional ratings of the participants were recorded and stored separately at the end of each game.

IV. PHYSIOLOGICAL FEATURE MATRIX CONSTRUCTION

In this study, all pre-processing, windowing and data analyses have been implemented using MATLAB software (version R2015b).

A. Pre-Processing

A 5th order Butterworth band-pass filter was applied to the raw EEG signals, whilst the lower-band was set to 4 Hz and the upper-band was fixed at 45 Hz (also employed by [14]-[17], [20]). This was due to the fact that eye-blinking artefacts are mainly observed in frequencies lower than 4 Hz, as people rarely blink more than four times a second. Thus, selecting 4 Hz as the lower-band cut-off frequency attenuates all blinking effects, present in the raw EEG signals. Moreover, the brain's high frequency rhythms (Gamma range) can be observed between 30 Hz and 45 Hz [33]. Therefore, selecting 45 Hz as the upper-band cut-off frequency attenuates all higher unwanted frequencies.

B. Windowing

To extract the affective features, a portion (called a Window) of the corresponding raw physiological signal is extracted and analysed. Any affective feature, extracted from this portion of the physiological signal, has to be able to be confidently tagged by a specific emotional experience. The emotionally labelled affective features, extracted from this period, are employed as a single observation, within the affective database, for the emotion recognition training process. The duration of each window could be either shorter than (as employed by [21], [23], [24]) or equal (as used by [20], [25]) to the duration of the each affective stimulus. On the other hand, the duration of windows (either shorter or equal to stimuli duration) can be either a fixed value (such as 2, 4, 10, etc. seconds), regardless of the affective experience duration; or a relative value (such as 10%, 20%, etc.) to be calculated, independently, according to the duration of the participants' affective experience. In this study, 28 arbitrary window lengths (17 fixed and 11 relative values) have been selected to be used and compared in the classification process. All relative durations would be shorter than the stimuli duration, except the 100% window length, which behave as a window with the entire stimuli duration. Therefore, in this study, both windowing techniques, with durations equal to and shorter than stimuli length, have been implemented and evaluated.

To perform spectral analysis on the signals, we employed a Fast Fourier Transform (FFT) technique. One of the hypotheses of the FFT analysis technique is the periodicity of the target signal [34]. However, the recorded physiological signals are not periodic waves. Applying FFT on non-periodic signals would cause a Spectral Leakage effect, which results in non-zero spectral powers in high frequencies, which may not belong to the original signal [35]. To eliminate this effect, weighting window functions can be applied to the signal before FFT analysis takes place [35]. In this study, Hamming windows have been employed, for the window weighting process. By applying non-overlapped windows, almost 50% of

the signal values, passed through the Hamming windows, would be attenuated by 50%. Consequently, this significant attenuation could result in considerable database signal loss. To resolve this issue, overlapping windows are employed to share the attenuated signal points with other windows. To avoid any maximum signals attenuation larger than 5%, the Hamming windows have to share (overlap) 80% of the signals, in the windowing process.

C. Psychophysiological Features

1. EEG Features

The Theta, Slow-Alpha, Alpha, Beta and Gamma frequency rhythms [33] have been extracted from all 14 single and seven symmetric-paired channels. Moreover, the Asymmetric power ratios [40], [41], for both Slow-Alpha and Alpha rhythms, have been extracted. Furthermore, the left frontal (AF3, F3, F7 and FC5), right frontal (AF4, F4, F8 and FC6), left parietal (P7 and O1), right parietal (P8 and O2), frontal (AF3, AF4, F3, F4, F7, F8, FC5 and FC6), parietal (P7, P8, O1 and O2) and overall EEGw3, have been calculated. In addition, the Alpha-Beta Ratio measurement, presented in (1), has been implemented in this study. According to [33], Alpha waves can indicate a relaxed awareness, without any attention or concentration, whereas Beta waves can be associated to active thinking, active attention or solving concrete problems. Therefore, this ratio can indicate an "attention measure" in a location of the brain (a large Alpha-Beta Ratio indicates high alpha activities and lower beta activations, signifying lower attention and concentration).

Alpha-Beta Ratio Equation;

$$\text{Alpha - Beta Ratio} = \frac{\text{Beta Power}}{\text{Alpha Power}} \quad (1)$$

The Alpha-Beta Ratio has been extracted from all 14 single and seven symmetric-paired channels. Therefore in total, 147 EEG features have been extracted from each window, retrieved from the raw EEG signals.

2. GSR and Heart Rate Features

The mean, minimum, maximum, standard deviation, mean of the peaks, mean of the first derivative, mean of the positive values of the first derivation, mean of the negative values of the first derivation, mean of the first derivative peaks and fluctuation frequency (The fluctuation frequency signifies the number of times the signal changes direction – i.e. increase to decrease and vice versa) have been extracted from both GSR and heart rate raw signals. Moreover, the GSR low frequency power (0 Hz to 2.4 Hz [16], [17]), heart rate medium (0.04 Hz to 0.15 Hz [16], [17]) and high (0.15 Hz to 0.5 Hz [16], [17]) frequency power and heart rate spectral power ratio ($\frac{\text{Medium Frequency Spectral Power}}{\text{High Frequency Spectral Power}}$) were also extracted from each retrieved window. Therefore, in total, 24 features were extracted from the raw GSR and heart rate signals in each window.

3. Participant Features

In total, three features, related to the participant, have been extracted, in each window. These are the **gender** (male vs. female), **hand preference** (right vs. left handed) and **age** (four classes: 12-18, 18-24, 24-30 and 30-40 years old), each of which has been recorded within the features matrix.

4. Affective Tagging

As the self-assessments are conducted at the end of each game, rather than continuously during the gameplay (Section III B), the following hypothesis has been presented in this study. First, we divided the emotional experience of the participants, during a single game, into two affective periods:

1. **'Emotion Build-Up' Period:** This period occurs during the first part of each game. Within this period, the emotional experience of the participant can be unpredictable, as it can be representative of a residual state from a previous game or some other pre-cognitive state.
2. **'Emotion Persistence' Period:** This period occurs during the last part of each game. Within this period, the emotional experience of the participant has been influenced by the current game, and can be (reasonably) confidently labelled by an Affective Cluster or Emotion Label. This means that all emotional experience variations within this period are considered as minimal. This also means that the affective experience of the participants within this period is always close to the average affective label (Relaxed, Content, Happy, Excited, Angry, Afraid, Sad and Bored) and cluster (PVLAPD, PVHPAPD, NVPAND and NVNAND). The cluster is determined according the dimensional ratings of the participants at the end of each game and the cluster boundaries, presented in [3]), reported by the participants at the end of the game.

Then, we hypothesised that the first 30% duration of each game constitutes the Emotion Build-Up Period, while the last 70% can be considered as the Emotion Persistence Period. Therefore, all windows, which have a centre, time-stamp within the first 30% period of the each game have been deleted from the features matrix. Then, all windows, which have a centre time-stamp within the last 70% period of the each game, have been tagged by the Affective Cluster and Emotion Label reported by the participants at the end of that game.

5. Defective Data Removal

All windows exhibiting EEG signals with an average signal quality below "fair" (according to the EMOTIVE EPOC signal quality classes) have been removed from the features matrix. Furthermore, all windows exhibiting infinity or NAN (Not-A-Number) values have also been removed from the features matrix.

V. FEATURE SELECTION

As it was discussed in Section IV C 174 features were extracted from all windows. To be able to perform emotion

classification, the dimension of the features matrix has to be reduced to a subspace. This subspace has fewer features (labelled Most Optimum Features throughout the paper), while they can adequately capture the essence of the data [30]. To perform the feature selection, the minimal-redundancy-maximal-relevance (mRMR) technique has been employed. Consider the features matrix of $F \in \mathbb{R}^{(N \times D)}$, while N is the number of observations and D is the number of features. The mRMR algorithm finds the most optimum subset $F_S \in \mathbb{R}^{N \times d}$, such that $d \ll D$, and F_S can optimally characterise F [36].

The mRMR algorithm employs Shannon's Entropy [37] to identify those features, which are mutually exclusive with respect to each other (minimal redundancy), whilst remaining mutually inclusive with respect to the classification clusters (maximal relevance – Affective Clusters or Emotion Labels in this study) [36]. To perform the analysis, the database has to be discretised prior to the Shannon's Entropy calculations. Therefore, all features were discretised according to three classes (-1, 0 and 1), with respect to the features' mean and standard deviation values (as implemented by [36]).

In the present study, 30 arbitrary values have been used as the number of required features ($d - 1$ to 30), each of which could be selected according to either Affective Clusters or Emotion Labels. Furthermore, the mRMR technique was capable of producing various lists of most optimum features, according to different windowing techniques employed in the features matrix construction process (28 different window lengths). This combination can create 840 different settings ($28 \times 30 = 840$), for classification according to either Affective Clusters or Emotion Labels.

VI. CLASSIFICATION AND AFFECTIVE RECOGNITION

In this study, the SVM [38] and KNN [39] classifiers have been employed to perform the classification process. To evaluate the performance of the SVM classifier, according to different settings, the Linear, 2nd Order Polynomial (Quadratic), 3rd Order Polynomial (Cubic) and Gaussian Kernel functions have been employed. Also 24 arbitrary Kernel Scales, for the Gaussian Kernel function, have been arbitrarily selected and evaluated in the cross-validation process. Also 30 different arbitrary K values, for the KNN classifier, have been implemented and evaluated in the cross-validation process (1 to 30). All classifications and cross-validations have been implemented within MATLAB software (version R2015b), using the Statistics and Machine Learning Toolbox.

A. Number of Features Evaluation

Fig. 2 presents the performance of the classifiers with respect to different number of features, according to Affective Clusters and Emotion Labels. The scattered dots in Fig. 2 (and also Fig. 4) present different classifiers with various settings. For example, if the classifier employs five features for the classification, different window lengths (28 different window lengths) and classifier settings (different K -value in KNN, etc.) can result in various accuracies (all scattered dots presented in a vertical manner for five features). However, as

the best performing classifier in each setting has to be selected, the setting, which generates the maximum classification accuracy is identified and highlighted (e.g. the line highlighting the maximum values in Fig. 2 and also Fig. 4). As it can be obtained by Fig. 2, the performance pattern of each classification technique, with respect to the number of features, are similar by employing either Affective Clusters or Emotion Labels. The accuracy of both KNN and SVM classifiers, with respect to the number of employed features, follows a sigmoid function pattern ($sigm(x) = \frac{e^x}{1+e^x}$, the term

“sigmoid” means S-shaped [42]). This means that their accuracies have increased by employing more features, with saturation occurring around 98%. As it can be seen in the graphs, the accuracy of classifiers has increased around 0.6%, by increasing the number of features from 20 to 30. By increasing the number of features, the complexity of the classifier grows, which consequently increases the classifier’s processing and timing expense. Therefore, we decided to not to employ more than 30 features in the classification process.

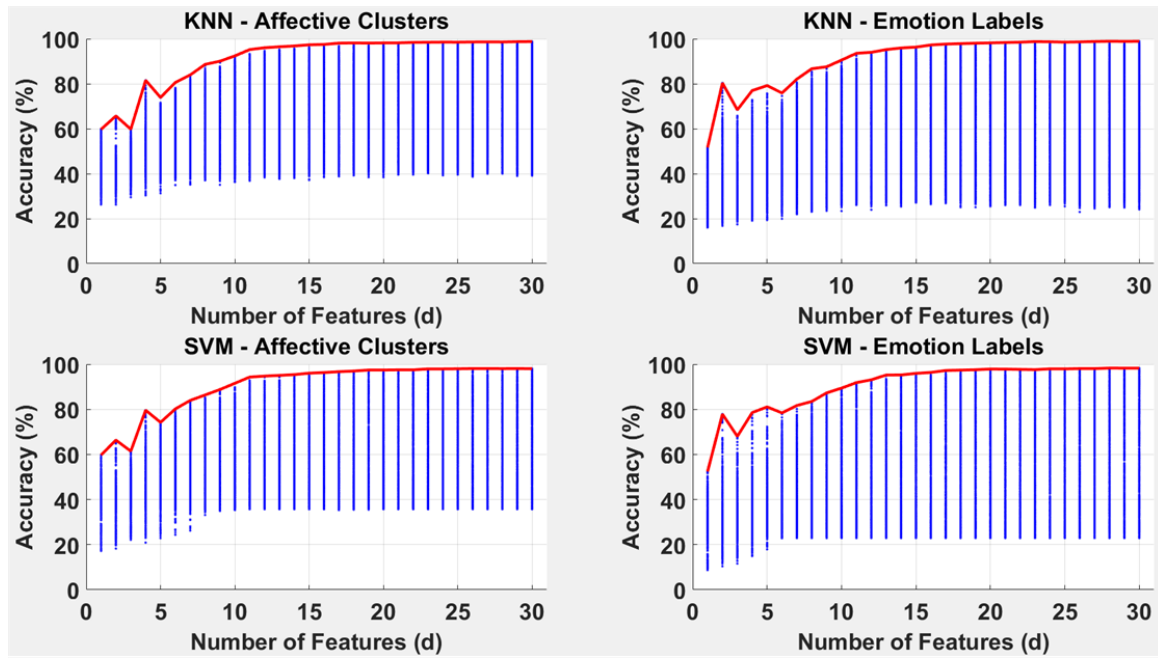


Fig. 2 KNN and SVM Classifiers Performance vs. Number of Features, According to Affective Clusters and Emotion Labels

B. Classification Settings Evaluation

Fig. 4 presents the performance of KNN and SVM classifiers, with respect to the corresponding classification settings, according to Affective Clusters and Emotion Labels. As it can be obtained by the figure, the performance of the KNN classifier is slightly attenuated, whilst “K” is increased (in classification according to both Affective Clusters and Emotion Labels). This means that the KNN classifier performs better when considering fewer neighbours in the affective space, in its attempts to classify the affective features. According to this analysis the 1st Nearest Neighbour (K=1) has the highest accuracy, compared to other “K” values. Also as illustrated by the graph, the performance of the SVM classifier is boosted when a higher order non-linear Kernel function is employed. The Gaussian Kernel function with relatively large kernel scales (either 2 or 3) performed better than the Linear and Quadratic Kernel functions. Although the Cubic Kernel performance was very similar to the Gaussian function, the best performing classifiers (Section VII) employed the Gaussian Kernel.

VII. DISCUSSION

To be able to compare the performance of all classification techniques, the best performing classifier setting (e.g. K value in KNN, etc.), for each window length, has been identified. As a result, 28 settings for each classification technique (KNN and SVM according to both Affective Clusters and Emotion Labels) have been identified. Fig. 3 presents the best classification accuracy, for each classifier, in each window length. The horizontal axis of the figure presents 28 different window lengths; 17 Fixed (left side of the vertical dashed line) and 11 Relative (right side of the vertical dashed line). An Analysis of Variance (ANOVA – Classifiers accuracy is considered as the dependent variables, while different classifiers, different windowing lengths and Affective Clusters vs. Emotion Labels classification technique as the independent parameters) showed that the performance of the classifiers, in categorising the emotions into either Affective Clusters or Emotion Labels is not statistically different (PClusters = 0.569). Also, the same analysis highlighted that the different windowing techniques (fixed vs. relative) is not a significant factor in changing the classifications performances (PWindowing = 0.691). However, the performances of KNN

and SVM classifiers are significantly different in terms of their classification accuracy ($P_{\text{classification}} < 0.001$). On average, KNN (97.01% ($\pm 1.3\%$)) mean accuracy across different windowing techniques, Affective Clusters and Emotion

Labels) outperformed the SVM algorithm (92.84% ($\pm 3.67\%$)) mean accuracy across different windowing, Affective Clusters and Emotion Labels) with around 4%.

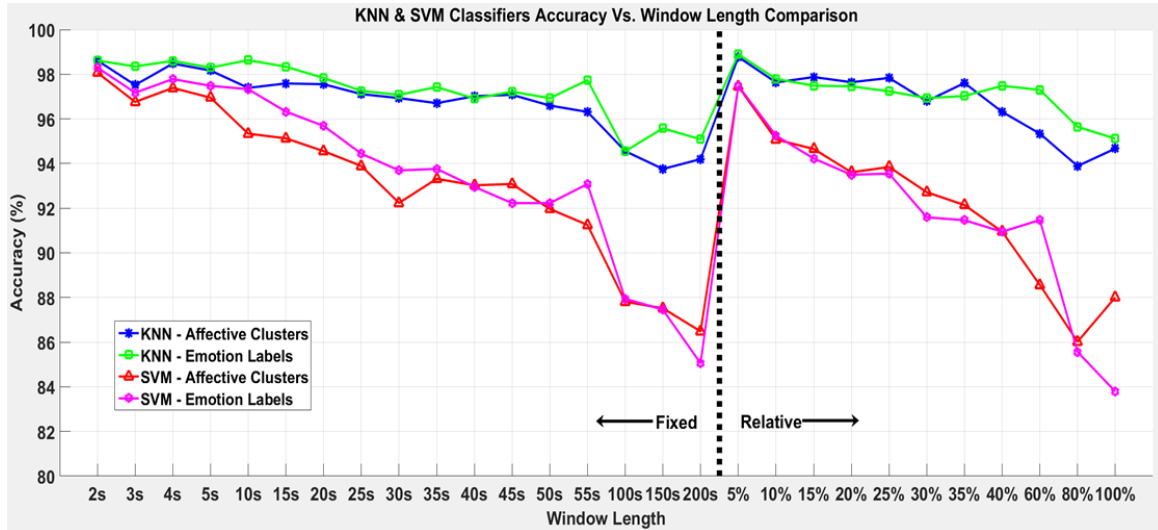


Fig. 3 KNN and SVM Classification Accuracy Comparison, vs. Window Lengths – According to Affective Clusters and Emotion Labels – The Horizontal Axes Presents 28 Window Lengths; 17 Fixed (Left Side of the Vertical Dashed Line) and 11 Relative (Right Side of the Vertical Dashed Line)

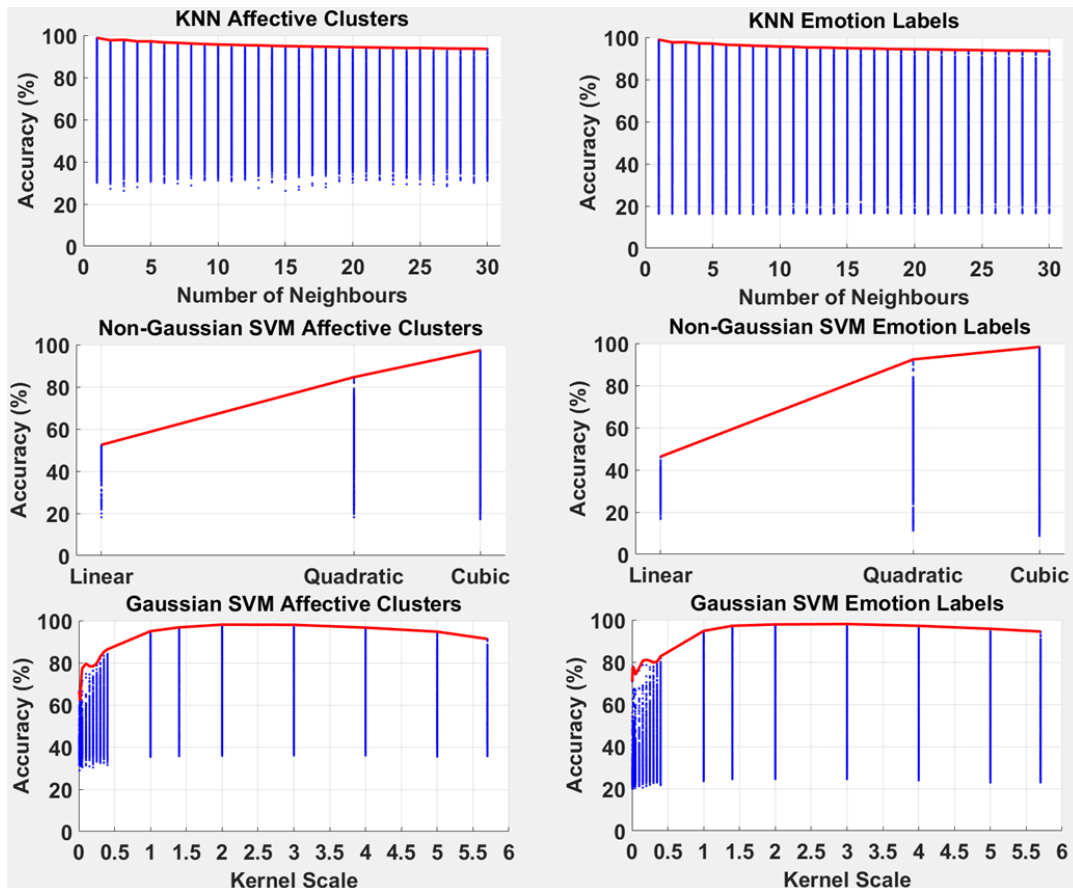


Fig. 4 KNN and SVM Classifiers' Settings vs. Accuracy, According to Affective Clusters and Emotion Labels

VIII. CONCLUSION

This paper demonstrated the phases of designing, conceptualisation and evaluation of an affective computing system, implemented in virtual reality. The findings of this study suggested that the physiological signals could be employed to classify emotional experiences. By assessing the performance of 28 different windowing techniques, we concluded that there is no difference in employing either relative or fixed windowing techniques. Therefore, as the relative windowing technique cannot be implemented in real-time applications (as the duration of the stimuli cannot be determined until the end of the VR session), the fixed windowing technique could be a more appropriate and credible choice to be adopted for real-time applications. However, the analysis suggested that the shorter window length could perform better in the classification process.

The final motivation of this research is to implement the designed affective recognition system, into an Adaptive Virtual Reality (Adaptive VR) demonstration, capable of adapting its internal environment according to the human users' emotion. Such a development could have significant implications for the development of dynamic human-centred interface techniques, supporting efficient human-system communication styles in a wide range of real-world applications.

REFERENCES

- [1] Paul Cairns, Anna Cox, Nadia Berthouze, Samira Dhoparee, and Charlene Jennett, "Quantifying the experience of immersion in games," in Workshop on the Cognitive Science of Games and Gameplay, Vancouver, 2006.
- [2] Anton Nijholt, Bos, Danny Plass-Oude, and Boris Reuderink, "Turning shortcomings into challenges: Brain-computer interfaces for games," *Entertainment Computing* 1, vol. 1, no. 2, pp. 85-94, 2009.
- [3] Mohammadhossein Moghimi, Robert J. Stone, Pia Rotshtein, and Neil Cooke, "Influencing Human Affective Responses to Dynamic Virtual Environments," *Teleoperators and Virtual Environments*, vol. 25, no. 2, November 2016.
- [4] Gunnar Ahlberg et al., "Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies," *The American Journal of Surgery*, vol. 193, no. 6, pp. 797-804, 2007.
- [5] Michael Zyda, "From visual simulation to virtual reality to games," *Computer*, vol. 38, no. 9, pp. 25-32, 2005.
- [6] Neal E. Seymour et al., "Virtual Reality Training Improves Operating Room Performance," *Annals of Surgery*, vol. 236, no. 4, pp. 458-464, 2002.
- [7] Nicole E Mahrer and Jeffrey I. Gold, "The Use of Virtual Reality for Pain Control: A Review," *Current pain and headache reports*, vol. 13, no. 2, pp. 100-109, 2009.
- [8] Hunter G. Hoffman, Jason N. Doctor, David R. Patterson, Gretchen J. Carrougher, and Thomas A. Furness, "Virtual reality as an adjunctive pain control during burn wound care in adolescent patients," *Pain*, vol. 18, no. 2, pp. 305-309, 2000.
- [9] A A Rizzo et al., "Virtual environments for the assessment of attention and memory processes: the virtual classroom and office," *Virtual Reality*, pp. 3-12, 2002.
- [10] David Jack et al., "Virtual Reality-Enhanced Stroke Rehabilitation David," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 9, no. 3, pp. 308-318, 2001.
- [11] Thomas D. Parsons and Albert A. Rizzo, "Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 39, no. 3, pp. 250-261, 2008.

- [12] JoAnn Difede et al., "Virtual Reality Exposure Therapy for the Treatment of Posttraumatic Stress Disorder Following," *The Journal of clinical psychiatry*, vol. 68, no. 11, pp. 1639-1647, 2007.
- [13] Marian Joels, Zhenwei Pu, Olof Wiegert, Melly S. Oitzl, and Harm J. Krugers, "Learning under stress: how does it work?," *Trends in Cognitive Sciences*, vol. 10, no. 4, pp. 152-158, 2006.
- [14] Mohammad Soleymani, Maja Pantic, and Thierry Pun, "Multimodal emotion recognition in response to videos," *IEEE Transactions on Affective Computing*, vol. 3, no. 2, pp. 211-223, April 2012.
- [15] Mohammad Soleymani, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17-28, May 2015.
- [16] Mohammad Soleymani, Sander Koelstra, Ioannis Patras, and Thierry Pun, "Continuous emotion detection in response to music videos," in 2011 IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), Santa Barbara, 2011, pp. 803-808.
- [17] Sander Koelstra et al., "DEAP: a Database for Emotion Analysis Using Physiological Signals," *Affective Computing, IEE Transactions*, vol. 3, no. 1, pp. 18-31, January 2012.
- [18] Christos A. Frantzidis et al., "On the Classification of Emotional Biosignals Evoked While Viewing Affective Pictures: An Integrated Data-Mining-Based Approach for Healthcare Applications," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 309-318, March 2010.
- [19] Peter J. Lang, Mark K. Greenwald, Margaret M. Bradley, and Alfons O. Hamm, "Looking at pictures: affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261-73, May 1993.
- [20] Kazuhiko Takahashi and Akinori Tsukaguchi, "Remarks on Emotion Recognition from Multi-Modal Bio-Potential Signals," in *Systems, Man and Cybernetics*, 2003. IEEE International Conference, vol. 2, Yamaguchi Univ., Japan, 2003, pp. 1654-1659.
- [21] Mimma Nardelli, Gaetano Valenza, Alberto Greco, Antonio Lanata, and Enzo Pasquale Scilingo, "Recognizing emotions induced by affective sounds through heart rate variability," *IEEE Transactions on Affective Computing*, vol. 6, no. 4, pp. 385-394, October 2015.
- [22] Herbon Antje, Christian Peter, Lydia Markert, Elke Van Der Meer, and Jörg Voskamp, "Emotion studies in HCI-a new approach," in *HCI International Conference*, vol. 1, Las Vegas, 2005.
- [23] Christos D. Katsis, Nikolaos Katertsidis, George Ganiatsas, and Dimitrios I. Fotiadis, "Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 38, no. 3, pp. 502-512, May 2008.
- [24] Dongrui Wu et al., "Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 109-118, July 2010.
- [25] Alejandro Rodríguez, Beatriz Rey, Miriam Clemente, Maja Wrzesien, and Mariano Alcañiz, "Expert Systems with Applications Assessing brain activations associated with emotional regulation during virtual reality mood induction procedures," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1699-1709, February 2015.
- [26] James A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161-1178, 1980.
- [27] Albert Mehrabian, "A Semantic Space For Nonverbal Behaviour," *Consulting and Clinical Psychology*, vol. 35, no. 2, pp. 248-257, October 1970.
- [28] Paul Ekman and Wallace V. Friesen, *Unmasking The Face*. Los Altos: ISHK, 2003, ISBN: 0-13-938175-9.
- [29] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 6, pp. 1052-1063, November 2011.
- [30] Kevin P. Murphy, "Introduction," in *Machine Learning: A Probabilistic Perspective*.: MIT Press, 2012, vol. 1, pp. 2-12.
- [31] Michael Steinbach, Vipin Kumar Pang-Ning Tan, "Introduction to Data Mining," in *Introduction to Data Mining*.: Addison-Wesley, 2005, pp. 65-73.
- [32] T. Tamura, H. Miike K. Nakajima, "Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique," *Medical Engineering & Physics*, vol. 18, no. 5, pp. 365-372, 1996.

- [33] Saeid Sanei and Jonathon Chambers, "Brain Rhythms," in EEG Signal Processing. West Sussex: John Wiley & Sons, 2009, pp. 10-13.
- [34] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, "Chapter 12 - Fast Fourier Transform," in Numerical recipes in Fortran (The art of scientific computing):. Cambridge University Press, 1992, vol. 1, pp. 490-529.
- [35] Fredric J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," Proceedings of the IEEE, vol. 68, no. 1, pp. 51-83, January 1978.
- [36] Hanchuan Peng, Fuhui Long, and Chris Ding, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226-1238, August 2005.
- [37] Kevin P. Murphy, "Entropy," in Machine Learning: A Probabilistic Perspective: MIT Press, 2012, vol. 1, p. 57.
- [38] Kevin P. Murphy, "Support Vector Machines," in Machine Learning A Probabilistic Perspective: MIT Press, 2012, vol. 1, pp. 498-507.
- [39] Kevin P. Murphy, "A Simple non-Parametric Classifier: K-Nearest Neighbors," in Machine Learning A Probabilistic Perspective.: MIT Press, 2012, vol. 1, pp. 16-18.
- [40] M. Murugappan et al., "Time-Frequency Analysis of EEG Signals for Human Emotion Detection," in Springer-Verlag, Berlin, 2008, pp. 262-265.
- [41] M. Rizon, M. Murugappan, R. Nagarajan, and S. Yaacob, "Asymmetric Ratio and FCM based Salient Channel Selection for Human Emotion Detection Using EEG," WSEAS Transactions on Signal Processing, vol. 4, no. 10, pp. 596-603, October 2008.
- [42] Kevin P. Murphy, "Logistic Regression," in Machine Learning a Probabilistic Perspective: MIT Press, 2012, vol. 1, pp. 21-22.