Affective Virtual Environments: A Psychophysiological HCI System Concept

By
Mohammadhossein Moghimi

Thesis
Submitted to the University of Birmingham
For the Degree of
Doctor of Philosophy (PhD)

Department of Electronic, Electrical and Systems Engineering
2017

Acknowledgments

Firstly, I want to thank my supervisors, Prof. Robert Stone and Dr. Pia Rotshtein for their endless help, advice and support, not only during my study, but also throughout my personal difficult times. Their efforts have most certainly transformed my abilities, from a shallow and simple-minded student to a cautious and fairly experienced researcher.

Secondly, I want to thank Professor Peter Gardner, for assisting me in funding the equipment, I needed for the experiments, conducted in this study. Also I want to thank Dr. Sara Wade for helping me in this study. I want to thank all of my participants as well, who made this research possible by patiently contributing to my experiments.

Most importantly, I want to thank the love of my life, Fatemeh Donyamali, for her unconditional affections and support during my study. Having her by my side, throughout all ups and downs and difficult times, was undeniably the most important motivation, which have brought smiles back to my face.

Last but not least, I want to thank my mom and dad, from the bottom of my heart, for always being there. There is no doubt that without their support I would have never been able to be the person I am at the moment. Although, words may never show my true level of gratitude, words are the only way to express how grateful I am. Finally, I want to thank my brothers, Mohammadhasan and Mohammadali, for their kindness and patience during my study.

To my father…

Declaration

I, Mohammadhossein Moghimi, hereby declare that, except where specific references are made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other Universities. This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration, except where specifically indicated in the text.

1. Part of this study has been presented in the "7th Computer Science & Electronic Engineering Conference", in 2015 (CEEC2015). A summary of the study has been published in *IEEE Explore*.
2. Chapter 3 and part of Chapter 2 are based on pre-printed manuscripts, which have been published in the journal "PRESENCE: Teleoperators and Virtual Environments" (Issue 2 of Volume 25).
3. Chapter 5 is based on two manuscripts, submitted to the "IEEE Affective Computing" journal. The papers are have been accepted with minor corrections, at the time of submission.
4. A summary of Chapter 5 has also been accepted for presentation at the "19th International Conference on Affective Computing and Intelligent Interaction", in 2017, (ICACII 2017). A summary of the study will be published in the "International Science Index" proceedings.
5. Chapter 4 is based on two manuscripts, prepared for submission to the "Emotion - American Psychological Association" journal for peer review.
6. The author, together with his supervisors, is preparing a paper to publish all psychological and physiological recordings, as open-source datasets, to enable other researchers to investigate various aspects of the considerable amount of psychological and physiological behavioural data recorded during the execution of the present research.

List of Abbreviations

AugCog	Augmented Cognition
AC	Alternating Current
ANOVA	Analysis of Variance
BCI	Brain-Computer-Interaction/Interface
DC	Direct Current
DA	Discriminant Analysis
EMG	Electromyography
ECG	Electrocardiography
EEG	Electroencephalography
GSR	Galvanic Skin Response
HCI	Human-Computer-Interaction/Interface
IAPS	International Affective Picture System
IADS	International Affective Digital Sound
KNN	K-Nearest Neighbour
LDA	Linear Discriminant Analysis
MANOVA	Multivariate Analysis of Variance
mRMR	minimal-Redundancy-Maximal-Relevance
NVPAND	Negative Valence, Positive Arousal, Negative Dominance
NVNAND	Negative Valence, Negative Arousal, Negative Dominance
NVPA	Negative Valence, Positive Arousal
NVNA	Negative Valence, Negative Arousal
OP	Occurrence Probability
PPG	Photoplethysmography
PVLAPD	Positive Valence, Low Arousal, Positive Dominance
PVHPAPD	Positive Valence, High Positive Arousal, Positive Dominance
PVLA	Positive Valence, Low Arousal
PVHPA	Positive Valence, High Positive Arousal
QDA	Quadratic Discriminant Analysis
SVM	Support Vector Machine
VR	Virtual Reality
VE	Virtual Environment
2D	2-Dimensional
3D	3-Dimensional
SAM	Self-Assessment Manikin

Abstract

The recent "resurrection" of interest in Virtual Reality (VR), courtesy of new interface and gaming technologies 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. Some believe that true immersion may only ever be achieved through advanced brain-computer interfaces, but, until that day arrives, it is important to understand how it may be possible to measure and, indeed, influence human engagement and emotional connection with virtual worlds using psychophysiological techniques. This study aims to design an affective computing system, capable of classifying human emotional experiences within dynamic virtual environments. First, by conducting an experiment, containing 120 participants, the subjective distribution of 8 selected Emotion Labels (Relaxed, Content, Happy, Excited, Angry, Afraid, Sad and Bored) within a 3-dimensional affective space (Valence *vs*. Arousal *vs*. Dominance) have been investigated. The high and significant correlation level of Dominance and Valence axes suggested that a 2-dimensional "Circumplex" model (Valence *vs*. Arousal) is adequate to distribute the selected Emotion Labels and cover the entire affective space.

Based on the development of the Circumplex model (divided into 4 Affective Clusters), a controllable Affective Virtual Reality has been constructed (Affective VR – in fact a computer game capable of evoking multiple emotions within the users). Multiple variations of the content and interface parameters of this game (*incidents*) have been subjectively evaluated, using a 35-participant online survey. The affective power of more than 790 VR variations (called *sub-games*) have been estimated using the incidents' estimated affective powers, coupled with an approximation algorithm. As a result, 22 sub-games, which have the maximum probability of evoking a number of emotional experiences, in the part of the participants, have been selected. These selected sub-games have been subjectively evaluated, within a 68-participant experiment, with different gender, age and gaming experiences. According to the subjective evaluation results, it was concluded that the designed Affective VR is capable of manipulating participants' emotional experiences, by controlling a number of internal parameters, while keeping its overall

theme and interaction process the same. Moreover, it was concluded that the developed emotion forecasting technique could accurately predict the participants' emotional *colour* (type of the emotion), while underestimate their *intensity* (power of a particular emotion). Furthermore, it was highlighted that the participants with different age, gender and gaming experiences could have different emotional responses, when exposed to a single affective session.

The designed and evaluated Affective VR has been employed in a physiologically-based experiment, in which the EEG, GSR and heart rates of 30 participants have been recorded during exposure to the most powerful affective sub-games, identified in the evaluation stage. Only male and female gamers, aged between 18 and 30 years old, have been recruited in this experiment. This was due the fact that the analysis, conducted earlier, highlighted that these groups demonstrate a consistent emotional reaction to Affective VR variations, when compared to female non-gamers and any participant aged between 30 and 40 years old. The result of the physiological experiment was employed to construct various feature matrices (with distinct pre-processing settings), containing 743 psychophysiological features. By employing a feature selection technique, the dimension of the feature matrices have been reduced to only 30 of the *most optimal features* (those which have the maximum relevance to the classification clusters, with minimum redundancy), to be used in the classification process. Finally, four classification techniques (KNN, SVM, DA and Classification Tree) have been employed to perform the classification process. The performances of more than a quarter of million classifiers, under various classification and pre-processing settings, have been evaluated using a cross-validation technique. It was concluded that the KNN and SVM classifiers perform better when compared to Classification Tree and DA classifiers. The trained KNN and SVM classifiers achieved 97.01% (±1.28) and 92.84% (±3.69) mean cross-validation accuracies, across different pre-processing settings, respectively.

By conducting another physiological experiment, containing 15 participants (male and female gamers, all aged between 18 and 30, none participated in previous experiments), the performances of the trained classifiers have been evaluated within a new dataset. Unlike the cross-validation results, the classifiers have not been able to achieve appropriate classification accuracies, and perform either similar to or even worse than random classifiers. The results suggested that, although, there is a significant variation pattern between the recorded psychophysiological features, within different affective sessions, a significant individual difference can also be observed within the database. Although, it was concluded that the presented affective computing systems could be considered as subject-dependent classifiers; it was also highlighted that identification and appropriate adjustment of the sources of the individual differences are extremely important necessities, to increase the robustness of the trained classifier for more generalised performances.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1. Introduction

Virtual Reality (VR), and interactive 3D environments generally, have experienced a significant “comeback” of recent years, courtesy of developments in the gaming industry and the relentless demand for high-fidelity escapist experiences on the part of gamers and simulation users alike. Yet, despite many international initiatives involving the design and development of highly innovative and affordable Human-Computer Interaction (HCI) technologies in the quest for the ultimate “immersive” experience[1], some believe that true “immersion” may only ever be achieved through the use of advanced brain-computer interfaces (BCI) (Cairns et al., 2006). However, until that day arrives, it is important to understand how it may be possible to measure and, indeed, influence human engagement and emotional connectivity with virtual worlds using psychophysiological techniques.

1.1. Human Computer Interaction (HCI)

Introduction of automated calculators and programmable devices can be traced back to almost 4 centuries ago, but since then one of the most important issues facing designers of those systems has been that of their interface with their human users. Both hardware and software developments have accelerated to spread these calculation devices throughout the world and, today, computing devices, in one form or another, exist in almost every household. Also, this wide availability has modified their functionalities from some calculators to required daily equipment (Wagner et al., 2010). The introduction of graphical user interfaces (GUIs) can be considered as a turning point in the history of human-computer interaction, as the GUI enabled end users to simply see and understand complicated commands and processes in a rich visual and graphical environment. Outstanding developments and inventions of new technologies have provided much cheaper and more powerful hardware platforms, enabling engineers and software developers to design and implement much more sophisticated software to deal with the interaction between human and computer.

Moreover, the introduction of the graphical user interface has, in some part, helped to stimulate developments of the concept of “Virtual” or 3D Environments. These environments can be employed in several applications, such as education, training and entertainment. Although, these environments can be considered as a very powerful platform for presenting sophisticated information in a very intuitive format, techniques for allowing humans to present their own information and interaction styles to such a complex system can be considered as a more important and challenging issue. The term “interaction” indicates a two-way communication process, between the system and the user, which includes both input and output pathways (Figure 1). So far the interaction process has been mostly based on conventional methods, in that computer users typically use physical interaction devices – displays/monitors, keyboards, mice, joysticks, gamepads, and so on, to see, hear, act, sense stimuli and in some cases even talk to the system.

[1] Witness, for example, the wide range of visual displays, data inputs, haptic and other forms of devices available from “crowd-funding” platforms, such as Kickstarter and Indiegogo.

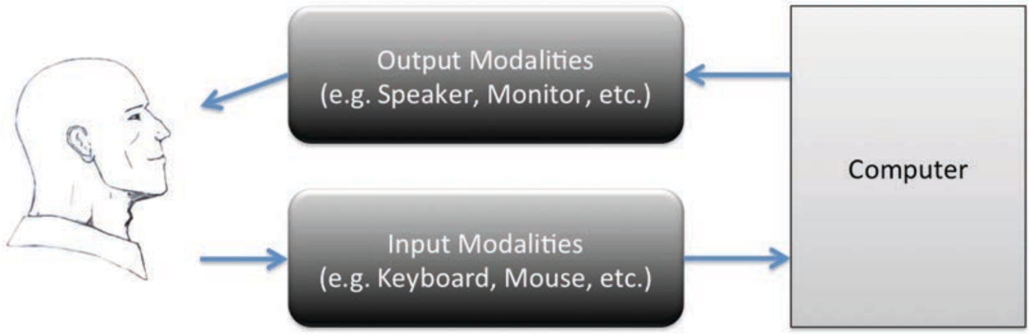


Figure 1 – Conventional HCI System

1.2. Virtual Reality (VR)

At its most fundamental, Virtual Reality (VR) refers to a "suite of technologies which permit intuitive interaction with real-time, three-dimensional databases" (Stone & Allardice, 1996) and is a form of simulation in which the end user interacts in real-time with multisensory, computer-generated databases (comprising predominantly, but not exclusively visual imagery; (Stone et al., 2017)). VR scenes can be presented to the human, using a variety of two-dimensional and three-dimensional (stereoscopic) display technologies, including conventional flat screens, head-mounted displays, smartphones and tablets, whole-wall displays and even room-sized enclosures formed of back-project screens (e.g. the "CAVE", or Cave Automatic Virtual Environment). Non-visual aspects of Virtual Environments ("VEs") include sound, haptic (delivering rudimentary sensations of touch and force), motion and olfaction (smell). Interaction with VEs (e.g. navigation and manipulation) can also be achieved using a wide variety of data input technologies, from whole- or part-body tracking systems, multi-axis hand-controllers, instrumented gloves (and suits), treadmills, speech recognition, eye tracking, and many more. VR is sometimes associated with the words "immersion" and "presence" (refer to Section 1.3), both of which tend to appear in marketing literature and online posts regarding the topic, but currently could be unachievable due to technical and human factors limitations with the hardware technologies. Data for generating VEs can be acquired from a number of sources, including commercial 3D modelling and image editing toolkits, computer-aided design (CAD) databases, specialised software tools for delivering complex real-world effects (i.e. those defined by the laws of physics – gravity, collision, water/fire/smoke particles, lighting, and so on) and even object or topographical 3D geometries generated using processed video, or from contact and non-contact sensors.

1.3. Immersion

The term "immersion" has most often been used to describe the multi-sensory experience of "presence" by individuals, whilst performing a task in VR. However, different researchers have suggested different definitions for this term (Brown & Cairns, 2004). As an illustration, Cairns et al. suggested that immersion could be defined as a feeling of being deeply engaged when people enter a make-believe world and feel as if it is real (Cairns et al., 2006). In 2004, Brown & Cairns suggested that immersion can be divided into three levels; engagement (during which the users invest time, effort and most importantly attention), engrossment (the time that the user's emotions are directly affected by the environment) and total immersion (when the users are detached from reality and the virtual world is, for them, all that matters). They claim that engagement and engrossment could be achieved much easier than a total level of immersion, believing instead that it could be achieved by overcoming other barriers. Such barriers include *Empathy*, as a "growth of attachment" to the environment, and *Atmosphere* as representing the VR's environmental realism. The authors also mentioned that total immersion could be difficult to achieve: "there are barriers to immersion from both the human and the system perspectives" (Brown & Cairns, 2004). Other researchers combine the immersive experience in virtual realities and 3D environments with the term *presence*, which is defined as "the extent to which a person's cognitive and perceptual systems are tricked into believing they are somewhere other than their physical location" (Patrick et al., 2000). Based on the variety of definitions of immersion evident in the literature, several discussions have been presented on the topic of how to evaluate immersive experiences. Many believe that true immersion might even be impossible to achieve with the present state of maturity in VR and gaming technologies. Others believe it could be achieved simply by defining the term more appropriately (Brown & Cairns, 2004).

1.4. Brain-Computer Interaction (BCI) and Psychophysiological Monitoring

Brain-Computer Interaction (BCI) and associated interface systems, which rely on psychophysiological forms of monitoring, are communication systems that detect messages and commands sent by individual to the external world, whilst they pass through the central and autonomic nervous systems (Wolpaw et al., 2002). These systems attempt to extract specific forms of electrical information from a number of specific psychophysiological signals, in an attempt to exploit them during the human-computer interaction process (Nijholt et al., 2009; Salvendy, 2006; Novak et al., 2012). These interfaces, as a whole, depend on two individual "processing modules" (the user's brain and physiological behaviours, and the computer), which operate independently, whilst remaining highly related. The user's brain and body generates several psychophysiological signals according to their own processes and cognitive states, and BCI systems use their own procedures to generate interpretations based on the user's brain and psychophysiological states. In this case the system requires a feedback system to enable both parties to adapt their "processing modules" and correct all misinterpretations. The brain, as an adaptive controller, would adapt itself with the system, while the BCI requires an adaptation process by employing a feedback system (Wolpaw et al., 2002). As an illustration, Leeb et al. performed a 4-month training process for both the BCI system and a paralysed user, to enable him to control a BCI-control-based wheel chair, within a virtual environment (Leeb et al., 2007).

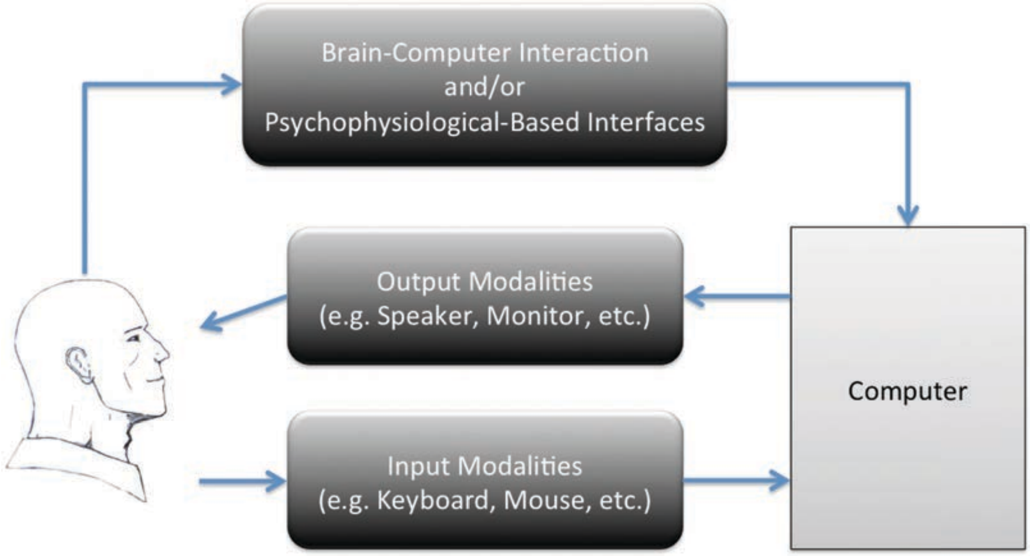


Figure 2 – HCI System with addition of BCI and/or Psychophysiological-Based Interface systems

BCI and psychophysiological-based interface systems also attempt to improve human-computer interaction and may potentially serve to increase the sense of immersion, by interfacing directly with the human nervous system (both the central and autonomic nervous systems), thus removing the artificial barriers to intuitive interaction afforded by conventional input-display techniques (Figure 2). These new interface channels could either be a replacement for the conventional input systems (e.g. translating imaginary movements into virtual actions), or have the potential to introduce a large number of new additional communication techniques in advanced HCI systems (e.g. improving levels of concentration, detecting emotional states, etc.), and may, therefore, be able to improve the interaction process considerably. The near-term goal of BCI and psychophysiological-based interfaces, as an extension to conventional HCI systems (as opposed to a replacement, which is a longer-term aspiration), would be to translate human thoughts, emotions, imaginary movement and actions by direct connection to the human central and autonomic nervous system, and to use this information as a new modality channel for the HCI systems (Nijholt et al., 2009). To do this, one or several psychophysiological signals have to be captured and interpreted to one or more operator functional states or "OFSs" to be used in the computer to either execute an action or change its internal states. (Salvendy, 2006).

1.5. Designing BCI and Psychophysiological-Based Interfaces

1.5.1. Psychophysiological Signal Recordings

One of the most important steps, in designing BCI and psychophysiological-based interface systems is the identification of specific, task-relevant psychophysiological signals, which carry particular information about the brain's cognitive states. Based on the system requirements and functionalities, a number of specific cognitive *features*[2] are required to be extracted, for further analysis and manipulations. These features should be extracted from psychophysiological signals captured from the human body; either directly from the brain (central nervous system) or other biologically related nerve systems (autonomic nervous system) (Salvendy, 2006; Nijholt et al., 2009). To date, a number of technologies have been implemented for recording various psychophysiological signals, for different purposes and requirements. A basic knowledge about the interactive system's functionalities, capabilities, constraints, required accuracy and speed need to be established, before selecting any of these technologies for the psychophysiological signal recording process (Salvendy, 2006; Nijholt et al., 2009). A number of these technologies are introduced below:

1. **EMG (Electromyography):** This is a technique developed to record and evaluate the electrical activities of the skeletal muscles. When the muscle cells are electrically or neurologically active, the generated electrical potential can be detected by this technique for further analysis.
2. **ECG (Electrocardiography):** This is a transthoracic[3] interpretation of the electrical activities of human heart over a period of time by using attached electrodes to the subject's chest (in close contact with human skin).
3. **PPG (Photoplethysmography):** During this recording a location of the skin is illuminated, and then the changes in light reflection are recorded. The DC (constant) component of the recorded signal is related to the bulk absorption of the skin tissue, while the AC (variations) component relates to the pulse pressure. The peak of the signal is also related to blood pressure under the target tissue.
4. **EEG (Electroencephalography):** The recording of electrical activities along the scalp of the human head. EEG measures voltage fluctuations resulting from ionic current occurring naturally within the neurons of the brain with respect to some stimulus.
5. **GSR (Galvanic Skin Response):** A method for measuring changes in the electrical conductance in the human skin. This information can be used to detect the level of stress as the physiological arousal level can control the sweating process.
6. **Skin Temperature Probe:** A device to measure heat changes on the human skin.
7. **Pupillary Response:** A physiological response, which varies the diameter of the pupil. This reaction could be triggered by a variety of causes, such as light exposure, attention, sexual stimulation, etc.

A commercial example of BCI-based technology, targeting the VR arena, was recently announced by the Neurable Company[4], specifically for integration within head-mounted displays. Their product claims to provide a stable neurological

---

[2] These features can be the statistical analysis of the physiological measurements (e.g. average heart rate), spectral analysis (e.g. specific frequency rhythms such as brain alpha waves), specific measurements (e.g. number of peaks), etc.
[3] Across or through the thoracic cavity or chest wall.
[4] http://www.neurable.com/

interface with virtual environments, although currently the system takes the form of a 7-electrode EEG interface integrated within the HTC Vive VR headset[5]. The system enables the users to control a virtual environment through brain-computer and/or conventional interfaces, such as motion joysticks, headset gyroscopes, and so on. Although this system represents one of the first commercial attempts to introduce BCI and psychophysiological-based interfaces to virtual environments and games, it is still the case that considerable improvements and revisions are required to enhance these interaction pathways.

1.5.2. Psychophysiological Signal Processing and Classification

After the stage of recording specific psychophysiological signals, the BCI and psychophysiological-based interface systems have to be able to undertake signal-processing techniques in order to extract related information and perform the required interpretations and/or classifications.

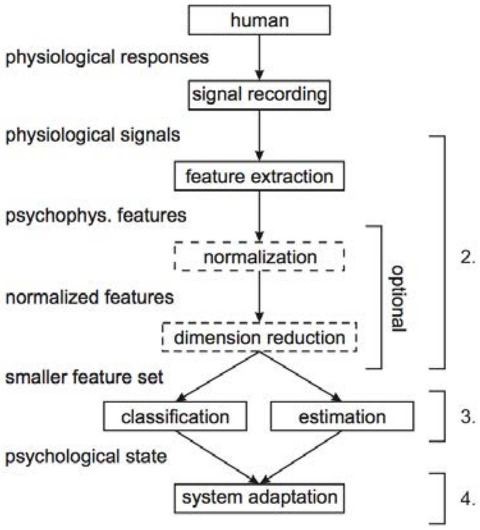


Figure 3 – The block diagram of general process of psychophysiological signal computing (Novak et al., 2012)

Figure 3 shows the process of interpretation of psychophysiological signals in a workflow diagram (Novak et al., 2012). The second section (highlighted by "2" in Figure 3) performs the prerequisites for the data fusion process (e.g. noise removal, normalisation, etc.), while the third part (highlighted by "3" in Figure 3) performs the data fusion process itself. The feature extraction block processes the incoming raw psychophysiological signals and extract some physiologically relevant features[2] to store them in a *feature matrix*, consisting of multiple valuable features. These features can have a high intra- and inter-subject variability as a result of age, gender, time of the day and several other factors (Novak et al., 2012). Normalisation helps to reduce this effect and make the feature matrix less variable for the data fusion process. If the number of psychophysiological responses were too many, the dimension of feature vector would grow exceedingly. The feature matrix will be used in the classification

[5] https://www.vive.com/sg/

and estimation processes. Thus, the process time would grow exponentially by increasing of the dimension of the feature vector. Moreover, a high feature matrix dimensionality could cause over-fitting in the classification or estimation process, which would attenuate the classification accuracy considerably. To solve this problem, a number of features could be removed, either by removing the features from the feature matrix, or by removing a particular sensor (Novak et al., 2012).

1.6. Affective Recognition/Computing

One of the sub-categories of research into BCI and psychophysiological-based interface systems is described as *affective computing*. During the process of affective computing, psychophysiological signals (central and/or autonomic nervous systems) from the users are recorded to enable the BCI system to extract data of relevance to their emotional and cognitive states. Such an 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, new techniques 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. (Koelstra et al., 2012; Frantzidis et al., 2010; Yazdani et al., 2009; Rizon et al., 2008; Murugappan et al., 2008; Katsis et al., 2008; Takahashi & Tsukaguchi, 2003)). Others are now beginning to focus on virtual realities and more interactive environments (e.g. (Parnandi et al., 2013; Wu et al., 2010; Antje et al., 2005)).

To date, many studies have been conducted to analyse the psychological and physiological responses of humans, in emotional experiences, evoked through videos (Soleymani et al., 2012; Soleymani et al., 2015), music videos (Koelstra et al., 2012; Soleymani et al., 2011), Images (Frantzidis et al., 2010; Lang et al., 1993), sound (Takahashi & Tsukaguchi, 2003; Nardelli et al., 2015) and real life scenarios (Katsis et al., 2008; Antje et al., 2005). Whereas less attention has been devoted to psychophysiological responses stimulated within Virtual Reality and Games (Wu et al., 2010; Rodríguez et al., 2015). To record the psychophysiological responses of the users, exposed to affective stimuli (images, video, etc.), various physiological recordings have been employed in the literature. To date, the most popular techniques with regard to the recording of human affective states have included EEG (Soleymani et al., 2012; Soleymani et al., 2015; Soleymani et al., 2011; Koelstra et al., 2012; Frantzidis et al., 2010; Takahashi & Tsukaguchi, 2003; Wu et al., 2010; Rodríguez et al., 2015), Galvanic Skin Response (GSR) (Soleymani et al., 2011; Koelstra et al., 2012; Frantzidis et al., 2010; Lang et al., 1993; Antje et al., 2005; Katsis et al., 2008; Wu et al., 2010) and Heart Rate (Soleymani et al., 2011; Koelstra et al., 2012; Lang et al., 1993; Takahashi & Tsukaguchi, 2003; Nardelli et al., 2015; Antje et al., 2005; Katsis et al., 2008; Wu et al., 2010). However, a small number of studies have also employed respiratory (breathing) rate, skin temperature, EMG and pupil diameter in order to classify affective states (Soleymani et al., 2012; Soleymani et al., 2011;

Koelstra et al., 2012; Wu et al., 2010) (for a more in depth literature review, refer to Section 5.1).

1.7. Affective Recognition in Virtual Realities

Turning briefly to the field of VR and the relevance of issues of affect, to date, researchers have studied the implementation of virtual realities in many different domains. As well as entertainment, virtual realities and their so-called "serious games" counterparts have been used for training purposes (Ahlberg et al., 2007; Zyda, 2005; Seymour et al., 2002), pain distraction (Mahrer & Gold, 2009; Hoffman et al., 2000; Hoffman et al., 2004), rehabilitation (Rizzo et al., 2002; Jack et al., 2001) and disorder therapy (Parsons & Rizzo, 2008; Difede et al., 2007; Rizzo et al., 2013; Kaganoff et al., 2012). The focus of all of 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 a pleasurable or dis-pleasurable condition) affect 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 (Joels et al., 2006). 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.

To date two different types of virtual environments have been employed in the study of affective recognition systems, to evoke different emotional experiences on the part of the users. Some studies have employed classical 2D "Retro Games"[6] such as "The Pong" (Liu et al., 2009), "Tetris" (Chanel et al., 2011) and "PAC-MAN" (Reuderink et al., 2013); whereas others have employed more graphical 3D environments, such as "Military-Class Humvee Driving" (Wu et al., 2010), "Race Car Driving" (Parnandi et al., 2013) and "Virtual Park" (Rodríguez et al., 2015) (See Figure 4).

In 2009, Lie et al modified the Pong game (Figure 4 – Section A) and employed an interactive version of the "Anagram Game" (a word reordering game) to evoke anxiety on the users (15 participants). They assessed the anxiety level of the participants through a subjective questionnaire. By employing EMG, GSR, heart rate and skin temperature signals, they achieved 88.9% accuracy (according to cross-validation), in classifying the physiological signals of the participants, into 3 classes (low, medium and high level of anxiety). By employing the same participants in another experiment, they managed to detect the participants' anxiety levels with 78% accuracy, while adapting the game difficulty accordingly (Liu et al., 2009).

In 2010, Wu et al employed a military Humvee simulator (Figure 4 – Section D) to fluctuate the arousal level of the users (18 participants). The participants were instructed to drive the Humvee in low and high threat environments, while performing

[6] https://en.wikipedia.org/wiki/Retrogaming

the Stroop Effect Test[7]. Instead of assessing the participants' emotional experience through subjective techniques, they estimated their arousal levels by employing the "Yerke-Dodson Law" (Yerkes & Dodson, 1908) and measuring the users' in-game performance (according to the number of correct answers in the Stroop Effect test). They employed EEG, GSR and heart rate signals to classify the participants' arousal levels, into 2 classes (high and low). Although they were able to achieve 84% accuracy in subject-dependent classifiers (tuned, trained and evaluated according to each individual), the classifiers performed randomly in subject-independent settings (trained on a group of participants and evaluated on the others) (Wu et al., 2010).

Figure 4 – Virtual Environments Employed Within Affective Computing Systems – **A)** "The Pong" (Liu et al., 2009) – **B)** "Tetris" (Chanel et al., 2011) – **C)** "PAC-MAN" (Reuderink et al., 2013) – **D)** "Military-Class Humvee Driving" (Wu et al., 2010) – **E)** "Race Car Driving" (Parnandi et al., 2013) – **F)** "Virtual Park" (Rodríguez et al., 2015)

In 2011, Chanel et al employed the Tetris game (Figure 4 – Section B) with three difficulty levels (easy, medium and hard) to study the emotional experience of a number of users (20 participants) under different game difficulties. Although they employed a 20-question subjective questionnaire to assess the participants' emotional experiences, the dimensionality of the subjective assessment was reduced to two principal components (using Principal Component Analysis – PCA); one related to pleasure and the other linked to arousal level. They concluded that the easy and hard

[7] reading colour word names with different coloured fonts; e.g. RED, Green, etc. https://en.wikipedia.org/wiki/Stroop_effect

difficulty levels are associated with less pleasurable experiences, when compared to the medium game difficulty. However, they discovered that the hard settings could evoke more arousing experiences, compared to medium levels; while medium settings could elicit more exciting experiences, when compared to easy levels. By measuring EEG, GSR, heart rate, breathing rate and skin temperature signals, they achieved 63% classification accuracy (according to cross-validation) in categorising the participants' emotional experiences, into 4 classes (combining low and high pleasure and arousal levels) (Chanel et al., 2011).

In 2013, Reuderink et al employed the PAC-MAN game (Figure 4 – Section C) to evoke frustration in 12 participants. They modified the game to have two different settings: (1) one with a normal hand controller and (2) another with faulty and unresponsive controller. They employed the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994) technique to assess the participants' emotional experiences, through Valence, Arousal and Dominance levels. They concluded that the setting with normal controller was associated with more pleasure and dominance level, when compared to the faulty controllers. However, they discovered that both settings were associated with similar arousal levels. By recording the EEG signals of the participants, when exposed to the games, they assessed the brain regional activities, according to the two emotional experiences. They concluded that there was a significant difference in the brain regional activities, between pleasurable and unpleasant, highly arousing, emotional experiences (Reuderink et al., 2013).

In 2013, Parnandi et al designed a racing car simulator (Figure 4 – Section E) with 3 difficulty levels, controlled according to three environmental and interactive parameters (weather, steering and the speed of the car). They employed 20 participants, and recorded their GSR activities while exposed to the racing car simulator, in two consecutive sessions; (1) open-loop and (2) close-loop. They hypothesised and concluded that the GSR peak frequencies (within 30 second windows) were related to variations in arousal levels (high arousing experiences generates higher GSR peak frequencies, and vice versa). Therefore in the open-loop session the average GSR peak frequency of each individual was recorded, to be employed as an emotional *set-point* (desired arousal level) in the close-loop session. The difficulty of the game was controlled in the close-loop session to maintain the GSR peak frequency of each individual around his/her corresponding set-point (Parnandi et al., 2013).

In 2015, Rodríguez designed a virtual park (Figure 4 – Section F) to evoke sadness and depression on 24 participants. The participants were instructed to navigate within the virtual park while depressing music was played at the background and a sad woman's voice guided them through the environment. At some points in the VR experience the participants encountered depressing images from the IAPS database (Lang et al., 2008), sad movie clips and depressing verbal statements appearing around various park locations. A significant decrease in positive emotion was recorded using a subjective measurement. Moreover by recording the EEG signals of the participants, before and during the VR experience, they concluded that the brain regional activity patterns of the participants, before and during the VR experience, are significantly different (Rodríguez et al., 2015).

1.8. Research Scope and Questions

As briefly discussed above, the majority of studies have investigated the design and implementation of affective computing systems within non-interactive environments. In the present study, we investigated the design and implementation of an affective computing system, within virtual and interactive environments. In this study the following questions have been investigated:

I. How the emotional experiences can be modelled, assessed and presented (Chapter 2)?
II. Is it possible to design an Affective VR, capable of manipulating the emotional states of the users (Chapters 3 and 4)?
III. Is it possible to design and train an affective computing system to recognise the emotional states of a number of users, using psychophysiological measurements (Chapter 5)?
IV. How well does the trained affective computing system (according to a number of participants) perform, when implemented within new datasets (new participants – Chapter 6)?

As presented above, there are a number of diverse research topics in this study. Each of these topics could require a unique and almost independent introduction and literature review. Combining the introductions and literature reviews of these topics, within a single chapter was felt by the author to be inappropriate, as it would contain a wide range of (almost) unrelated and independent introductory discussions. Therefore, in the present thesis, the introduction and literature review of every section (if required), is presented within itself. To summarise the scope of the research and the subsequent structure of this thesis, to answer the research questions:

In Chapter 2, we review the literature, defining emotions or so-called *Affect*, and present the two most famous affective models (categorical *vs*. dimensional), representing the emotional space (research question I). Then the possible techniques for evaluation and labelling of emotional experiences are discussed in detail. By conducting an experiment, containing 120 participants (Experiment 1), the relationship between two affective spaces (categorical *vs*. dimensional) is investigated. As a result, the dimensional affective space has been categorised into 4 Affective Clusters. These clusters have been employed further in the study, to manipulate participants' psychological and physiological reactions.

In Chapter 3, the design of an Affective Virtual Reality is described, the aim of which is to manipulate the participants' emotional experiences by changing a number of internal game parameters (research question II). By conducting a 35-participant online survey (the "Pre-Experiment" – Experiment 2), coupled with an approximation technique, the emotional effects of the Affective VR variations (called *sub-games*) have been estimated. By employing the sub-games' approximated emotional powers, the 22 most affective sub-games (those which have the highest possibility of evoking a particular emotion) have been identified and were presented to 68 participants (the "Preliminary Experiment" – Experiment 4) with different gender, age and gaming experiences. The participants were instructed to play the 22 sub-games and report their emotional experiences at the end of each sub-game.

In Chapter 4, the results of the Preliminary Experiment are assessed to evaluate the efficiency of the designed Affective VR (research question II). Moreover, the approximated affective powers of the sub-games are compared with the subjectively reported powers, to evaluate the performance of the designed emotional experience-forecasting algorithm. Furthermore, sources of individual differences between participants, which resulted in various emotional experiences within a single affective session, are analysed and highlighted. These sources have been appropriately adjusted to minimise emotional variations, among participants, when exposed to a single affective session.

In Chapter 5, first, a deep affective recognition literature review is conducted. The review highlights all of the design steps of various affective recognition systems (research question III). Then, by employing the designed Affective VR, a psychophysiological database, comprising EEG, GSR and heart rates of 30 participants, exposed to a number of sub-games, has been constructed (the "Primary Experiment" – Experiment 4). The psychophysiological database has been divided into multiple portions (called windows) with various lengths. Then 743 features have been extracted from the windowed psychophysiological database to generate the training feature matrices. By employing a feature selection technique, the dimension of the training feature matrices has been reduced to 30 features. Then the trimmed training feature matrices have been used in the classification process, using four classification algorithms. The performance of the trained classifiers has been assessed using the cross-validation technique, to reveal the best classification setting combination (pre-processing and classification settings).

In Chapter 6, the performance of the best performing classifiers (in the cross-validation step) is evaluated within another database containing the physiological responses of 15 new participants ("Evaluation Experiment" – Experiment 5– research question IV). The classifiers have been unable to perform appropriate classification, and, on average, performed either similar to or worse than random classifiers. By further analysis, the sources of this performance attenuation have been identified and reported.

Chapter 2

Models of Affect and Emotional Experience Assessment

Abstract – Defining the term "Emotion" or so-called "Affect" can be considered as one of the most important challenges of studies, dealing with psychological and/or physiological aspects of emotional experiences. So far many definitions and models have been presented in the literature, to represent the "Emotion Space". In this chapter the "Dimensional" (Valence *vs*. Arousal *vs*. Dominance) and "Categorical" (Emotion Labels, e.g. relaxed, stressed, etc.) models of affect are introduced and discussed in detail. Also, the possible techniques for assessing and recognising participants' emotional experiences have been introduced and considered. Moreover, by conducting a 120-participant experiment the relation between the categorical and dimensional affective models have been investigated. The results of the experiment suggested that the 2-dimensional affective space (Valence *vs*. Arousal – also known as Circumplex of Affect) is adequate in representing the selected "Emotion Labels", and the entire emotional space. The results highlighted that the entire 2-dimensional Circumplex of Affect can be categorised into four "Affective Clusters" (rather than four quadrants), some of which share two quadrants. This affective space clustering has been employed for further psychological and physiological investigations.

2. Models of Affect and Emotional Experience Assessment

2.1. Models of Affect

One of the most important challenges in the study of emotion is the definition one adopts. Psychologists have presented several interpretations and definitions for emotions (also known as *affects*), and there are almost as many definitions and models for affects and emotions as there are investigators (Bradley & Lang, 2006). The common ground amongst all of these emotional models is that emotions impact physiological, neurophysiological and cognitive responses, to enable the individual to react and perform adequately. In high-tempo, high-pressure contexts, for example, the heart rate changes, sweating occurs, the muscles tense, facial expressions change, and many other less overt physiological changes take place facilitating the so-called “fight or flight” reaction (Bradley & Lang, 2006). To classify and identify the emotional experiences, there are two distinct affective models; *Categorical* (*Qualitative*) and *Dimensional* (*Quantitative*).

1. **Categorical (Qualitative) Model:** 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 qualitative presentation of emotions, categorising them as surprise, fear, disgust, anger, happiness and sadness (Ekman & Friesen, 2003). Researchers have introduced several emotion sets, although there are some common strong emotions, which are present in majority of them. These strong emotions include anger, fear, disgust, excitement, happiness, sadness and boredom (Bradley & Lang, 2006).
2. **Dimensional (Quantitative) 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 (dimensions or axes) (Russell, 1980; Mehrabian, 1970). Mehrabian introduced three independent quantities: *Valence*, defining pleasure and displeasure; *Arousal* describing the excitation level; and *Dominance* identifying the level of control within a given situation. Russell, on the other hand, ignored Dominance, and created a 2-dimensional *Circumplex of Affect*. Mehrabian and Russell believed that representations of verbal labels of emotions, within either the 2D or 3D model, would differ between people with different cultures, especially those with different languages (Russell, 1980; Mehrabian, 1970). In 1980 Russell represented some of the most common English verbal labels within his Circumplex of Affect. Figure 5 presents a simplified version of the Russell’s Circumplex of Affect, occupying only eight Emotion Labels, which were distributed evenly within the model.

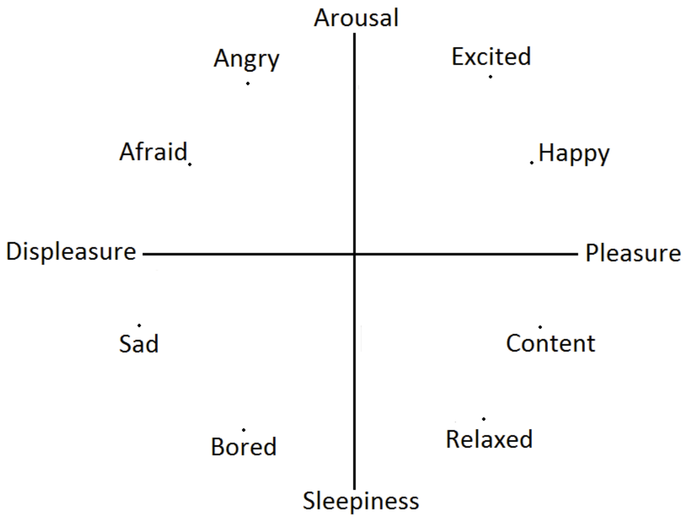


Figure 5 – Simplified Russell Circumplex of Affect for 8 English Verbal Labels of Emotions (Russell, 1980)

2.2. Emotional Experience Assessment

In the affective psychophysiological database construction process, each session, in which participants are exposed to an affective stimulus, has to be tagged by an emotional experience state, within an affective space (categorical and/or dimensional). This assessment has to be capable of reliably categorising the participants' emotional experience. So far studies have, in the main, employed *pre-affective hypothesis*, *expert* or *self*-assessments:

1. **Pre-affective hypothesis:** A hypothesis is presented in the study to tag the participants' experience with an affective state**.** As an illustration, Bailenson et al. employed an emotion recognition system as an affective tagging reference, which used the participants' facial gestures to classify the emotional experience. They hypothesised that the system classification based on facial gestures is reliable enough to be used as the reference point for the physiological affective recognition system's training process (Bailenson et al., 2008). As another example, Othman et al. hypothesised that the designed affective stimuli would evoke the target emotional experience on all participants (Othman et al., 2013).
2. **Expert assessments:** Here, a psychologist or human emotion expert is instructed to evaluate the participant's affective state, and to categorise it within an Affective Space (Katsis et al., 2008).
3. **Self-assessment:** There are two different techniques within this category. Firstly the minority of studies employed pre-evaluated affective datasets[8], and used the stimuli's average reported affective assessments as the reference point, instead of requesting the participants to evaluate their emotional experience after each stimulus (Frantzidis et al., 2010; Jenke et al., 2014). In the second technique, covering the majority of studies, participants were instructed to

[8] Like International Affective Picture System (IAPS) (Lang et al., 2008), Digital Sound (IADS) (Bradley & Lang, 1999) and affective video clip database (Baveye et al., 2013)

evaluate their emotional experience after each stimulus, rather than relying on pre-evaluated[8] assessments (if available) (Murugappan et al., 2008; Yazdani et al., 2009; Koelstra et al., 2012; Antje et al., 2005).

Employing a pre-affective hypothesis can cause unreliability within the database. For instance, a stimulus, hypothesised to evoke a particular emotion, could induce a completely different emotion on the part of the users. This mismatch between the hypothesised and evoked emotion could invalidate the classification process, as there is no emotional feedback to evaluate the reliability and effectiveness of the affective stimuli. Moreover, a psychologist or human emotion expert may be unavailable to conduct the emotional assessments on the participants who are exposed to affective stimuli. Therefore, employing self-assessment techniques, specifically for the purposes of emotion assessment, can be considered as the most convenient technique in the conduct of affective experiments. Indeed, and according to the findings of the literature review, the self-assessment technique has been employed by the majority of research studies (Section 5.1.7).

2.3. Affective Clustering – Subjective Experiment (*Experiment 1*)

2.3.1. Self-Assessment

According to a literature review conducted on 30 affective recognition studies, conducted since 2010 (refer to Section 5.1, for more details), it was highlighted that 40% of the studies employed a categorical model, whilst the other 60% used the dimensional system. This outcome left no room for concrete conclusions about the popularity of an affective space against the other. Therefore, in the present study, **both** dimensional and categorical affective spaces were investigated, in performing participants' self-assessments. Each axis was defined and presented to the participants as shown below:

1. **Valence:** How pleasurable the experience is. High positive values indicated more pleasure (e.g. the participant enjoyed it), and high negative values mean more displeasure (e.g. the participant did not enjoy it).
2. **Arousal:** How arousing the experience is. High positive values indicated greater arousal (e.g. excited, alert, stressful, etc.), and high negative values indicated minimal levels of arousal (e.g. relaxed, tired, bored, etc.).
3. **Dominance:** How much control the participant has during the experience. High positive values indicated higher control in the situation, and high negative values indicated lower control during the situation.

Furthermore, eight Emotion Labels were also selected, for categorical assessments (these were: Relaxed, Content, Happy, Excited, Angry, Afraid, Sad and Bored). As the original motivation of this study is to assess the emotional responses of participants within virtual environments and games, these eight Emotion Labels were selected, as they were assumed, in the judgment of the present authors, to be relevant to most VR experiences. These labels were also evenly distributed along the

Circumplex of Affect (Figure 5) to ensure that the entire 2-dimensional affective space is covered.

2.3.2. Participants and Method

To identify the **subjective arrangement** of selected Emotion Labels within the 3D affective space, 120 participants (mean age of 23.23 years, and a distribution of 53% male and 48% gamers – refer to Section 3.4.2, for gamer *vs*. non-gamer definition) completed a questionnaire to place the eight Emotion Labels (Relaxed, Content, Happy, Excited, Angry, Afraid, Sad and Bored) within the 3-dimensional affective space. These 120 participants have participated in all other experiments (Pre-Experiment survey (Section 2.3), Preliminary (Section 3.5), Primary (Section 5.2) and Evaluation (Section 6.1) Experiments), and completed a questionnaire at the end of the experiments. The questionnaire contained eight questions, each of which required the participants to locate one of the Emotion Labels within the 3D Affective Space (refer to Appendix A). The example given below presents one of the questions, assessing the "Relaxed" label. The participants were asked to choose one of the integer scalars, (arbitrarily) between -3 to +3, for each parameter. The study was reviewed and approved by the University of Birmingham's Ethical Review Committee (Ethical Reference Number: ERN_13-1157).

"*What value of these parameters (Valence, Arousal and Dominance) would describe the experience of "Being Relaxed" in virtual realities?*"

2.3.3. Results

Figure 6 presents the position of all eight Emotion Labels within the 3D affective space, according to their mean ratings (across all participants), in each affective axis. To simplify the 3-dimensional space presentation, three separate graphs have been obtained to demonstrate the Valence/Arousal, Dominance/Arousal and Valence/Dominance combinations. A Multivariate Analysis of Variance (MANOVA)[9] showed that the ratings of the eight Emotion Labels are significantly different ($P < 0.001$). By combining the Emotion Labels within each quadrant of the Russell's Circumplex of Affect (Figure 5) four Affective Clusters within the 3D Affective Space have been created (i.e. one to contain Relaxed and Content, another for Happy and Excited, etc.). A MANOVA[10] analysis showed that the ratings of the four Affective Clusters are significantly different ($P < 0.001$), as well. Table 1 presents the means and ranges of the ratings for both Emotion Labels and Affective Clusters, across all participants. As can be obtained from the ratings' means and percentiles, the majority of the ratings are within the Affective Clusters borders (for the ratings distribution plots, refer to Appendix B). Moreover, the significant rating difference of Emotion Labels and Affective Clusters (investigated through MANOVA analyses presented above) enable one to conclude that each Affective Cluster shares an insignificant and ignorable overlap with the other three clusters.

[9] Valence, Arousal and Dominance are considered as dependent variables and Emotion Labels as independent parameter.
[10] Valence, Arousal and Dominance are considered as dependent variables and Affective Clusters as independent parameter.

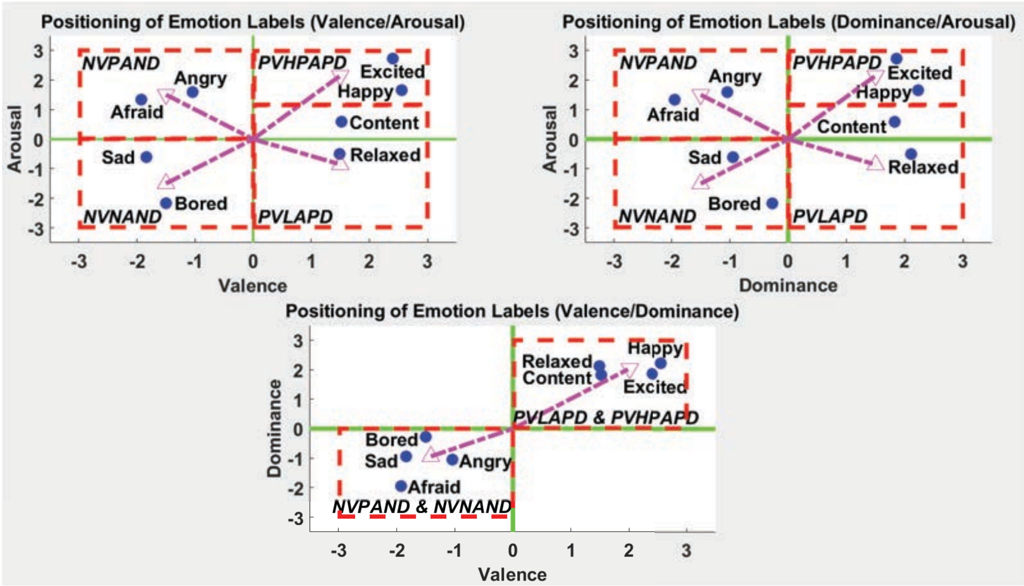


Figure 6 – Subjective Representation of 8 Emotion Labels, within 3D Affective Space – The Dashed Squares Present the Affective Clusters' Boundaries – The Dashed Arrows Present the Affective Clusters' Centre Vectors – Clusters are named as follows: (1) Positive Valence, Low Arousal, Positive Dominance *(PVLAPD)* – (2) Positive Valence, High Positive Arousal, Positive Dominance *(PVHPAPD)* – (3) Negative Valence, Positive Arousal, Negative Dominance *(NVPAND)* – (4) Negative Valence, Negative Arousal, Negative Dominance *(NVNAND)*

Table 1 – Mean Value of Valence, Arousal and Dominance Ratings of Emotion Labels and Affective Clusters, across all Participants' Survey Ratings – (A - B) Presents the (A) 25th and (B) 75th Percentiles, which Present the Range, and which Contain at Least 50% of the Ratings

Emotion	Mean (25th – 75th Percentiles)		Cluster	Mean (25th – 75th Percentiles)	
Relaxed	Valence	1.49 (1 – 2)	PVLAPD	Valence	1.50 (1 – 2)
	Arousal	-0.49 (-1.5 – 0)			
	Dominance	2.11 (1 – 3)		Arousal	0.04 (-1 – 1)
Content	Valence	1.52 (1 – 2)			
	Arousal	0.58 (0 – 1)		Dominance	1.97 (1 – 3)
	Dominance	1.82 (1 – 3)			
Happy	Valence	2.55 (2 – 3)	PVHPAPD	Valence	2.48 (2 – 3)
	Arousal	1.65 (1 – 2)			
	Dominance	2.23 (2 – 3)		Arousal	2.19 (2 – 3)
Excited	Valence	2.40 (2 – 3)			
	Arousal	2.72 (3 – 3)		Dominance	2.05 (2 – 3)
	Dominance	1.86 (1 – 3)			
Angry	Valence	-1.04 (-2 – 0)	NVPAND	Valence	-1.48 (-3 – -1)
	Arousal	1.59 (1 – 3)			
	Dominance	-1.05 (-3 – 0)		Arousal	1.45 (0.5 – 3)
Afraid	Valence	-1.92 (-3 – -2)			
	Arousal	1.32 (0 – 2.5)		Dominance	-1.50 (-3 – 0.5)
	Dominance	-1.95 (-3 – -2)			
Sad	Valence	-1.83 (-3 – -1)	NVNAND	Valence	-1.66 (-3 – -1)
	Arousal	-0.60 (-2 – 0)			
	Dominance	-0.95 (-2 – 0)		Arousal	-1.37 (-3 – 0)
Bored	Valence	-1.50 (-3 – 0)			
	Arousal	-2.15 (-3 – -2)		Dominance	-0.61 (-2 – 0)
	Dominance	-0.27 (-2 – 0)			

2.3.4. Axes Correlations

To assess the relationship between the axes of the dimensional affective space, the correlations between all possible pairs have been calculated; Valence *vs*. Arousal, Valence *vs*. Dominance and Arousal *vs*. Dominance. This analysis has been conducted on 4 independent datasets, recorded according to four independent experiments (survey, VR, image and sound).

1. **Survey:** This survey contains the subjective distribution ratings of Emotion Labels, scattered within the 3-dimensional affective space. By combining the ratings of 120 participants (Section 2.3.3), for eight Emotion Labels, 960 Valence *vs*. Arousal *vs*. Dominance ratings were recorded.
2. **Affective Virtual Reality (VR):** In this study the emotional experiences of 68, 30 and 15 participants, exposed to (respectively) 22, 10 and 5 Affective VR environments (Preliminary, Primary and Evaluation Experiments – Chapters 3, 5 and 6), have been self-assessed and recorded. The ratings have been combined and a database, with 1827 Valence *vs*. Arousal *vs*. Dominance ratings, was created.
3. **International Affective Picture System (IAPS):** This international database contains the subjective ratings of an affective picture set. The images' mean males and females ratings of the IAPS database (Lang et al., 2008) have been combined to create a database, with 2388 Valence *vs*. Arousal *vs*. Dominance ratings.
4. **International Affective Digital Sound (IADS):** This international database contains the subjective ratings of an affective sound set. The sounds' mean males and females ratings of the IADS database (Bradley & Lang, 1999) have been combined to create a database, with 334 Valence *vs*. Arousal *vs*. Dominance ratings.

Table 2 presents the correlation coefficients and significance levels of the axes correlations, within all four independent affective datasets. As it can be seen in the table, there is a high and significant correlation (Pearson's technique) between the Valence and Dominance ratings, whereas the correlation between other axes is considerably lower. Also this significant correlation is consistence in all four independent databases. This correlation renders the "Valence *vs*. Arousal" and "Dominance *vs*. Arousal" graphs almost identical (Figure 6). Furthermore, as it can be seen in Figure 6, the "Negative Valence, Positive Dominance and Positive/Negative Arousal" and "Positive Valence, Negative Dominance and Positive/Negative Arousal" octants[11] (4 octants out of 8) are completely empty (containing no emotion label), whilst the other four contain all eight Emotion Labels (see "Valence *vs*. Dominance" graph in Figure 6). From the Valence/Dominance correlation, one could conclude that the participants tend to report more pleasurable experiences (positive Valence) when they feel more control (positive Dominance) in the situation.

[11] An octant is one of the eight divisions of Euclidean three-dimensional coordinate system, defined based on the signs of the coordinates.

Similarly, they tend to report more displeasure (negative Valence) when they experience low control (negative Dominance) in the situation.

Table 2 – Axes Correlation Results, Conducted on Four Independent Databases

Database (Number of Samples)	Comparison	r-value	p-value
Survey (960)	Valence *vs*. Arousal	0.3376	<0.001
	Valence *vs*. Dominance	0.6910	<0.001
	Arousal *vs*. Dominance	0.1317	<0.001
Affective VR (1827)	Valence *vs*. Arousal	0.3060	<0.001
	Valence *vs*. Dominance	0.5595	<0.001
	Arousal *vs*. Dominance	-0.1011	<0.001
IAPS (2388)	Valence *vs*. Arousal	-0.2087	<0.001
	Valence *vs*. Dominance	0.8089	<0.001
	Arousal *vs*. Dominance	-0.5173	<0.001
IADS (334)	Valence *vs*. Arousal	-0.4062	<0.001
	Valence *vs*. Dominance	0.9282	<0.001
	Arousal *vs*. Dominance	-0.5027	<0.001

2.3.5. Discussion

The significant difference between the ratings of the Emotion Labels suggests that the dimensional presentation of these eight Emotion Labels are completely different from each other, and can be distinguished by the participants. As it can be seen in Valence/Arousal graph, in Figure 6, the labels followed the Russell's Circumplex order (Figure 5), whilst the position of "Relaxed" and "Content" were associated with slightly higher Arousal states than expected. It can be concluded that this reflects the fact that the ratings were undertaken in the context of the virtual reality and gaming experience; it is not unreasonable to suggest, or indeed expect that relaxing whilst playing a game can be more arousing than simply relaxing, for example, on a sofa.

Due to high and significant correlation between the Valence and Dominance axes (Section 2.3.4), in this study, it was decided to disregard the Dominance axis in all affective evaluations, and use the 2D Circumplex of Affect, originally presented by Russell (Russell, 1980) (Figure 5). Figure 7 presents the subjective positioning of eight Emotion Labels within the 2D (Valence *vs*. Arousal) Circumplex of Affect. Therefore, although, the four quadrants[12] in Russell's Circumplex of Affect could be employed to classify the emotional responses into four clusters (each of which contain two Emotion Labels – as employed by (Frantzidis et al., 2010)), the subjective arrangement of the labels within the evaluated Circumplex (Figure 7) categorises them into four **Affective Clusters**, rather than four quadrants, as the PVLA (Positive Valence, Low Arousal) and PVHPA (Positive Valence, High Positive Arousal)

[12] A quadrant is one of the four divisions of Euclidean two-dimensional coordinate system, defined based on the signs of the coordinates.

clusters share the positive valence and positive arousal quadrant. These four Affective Clusters can be employed for emotion recognition processes, categorising the affective responses into four clusters (PVLA, PVHPA, NVPA (Negative Valence, Positive Arousal) and NVNA (Negative Valence, Negative Arousal)). Also, as the combined ratings of the Emotion Labels, into Affective Clusters are significantly different, one could conclude that the four Affective Clusters are completely different from each other, and could be distinguished by the participants. Table 3 presents the Affective Clusters' centroids, plus defined upper and lower limits. Any rating within each cluster's limit could be tagged by the cluster's name. Therefore all dimensional affective ratings can be categorised, according to the clusters' limits, into four Affective Clusters.

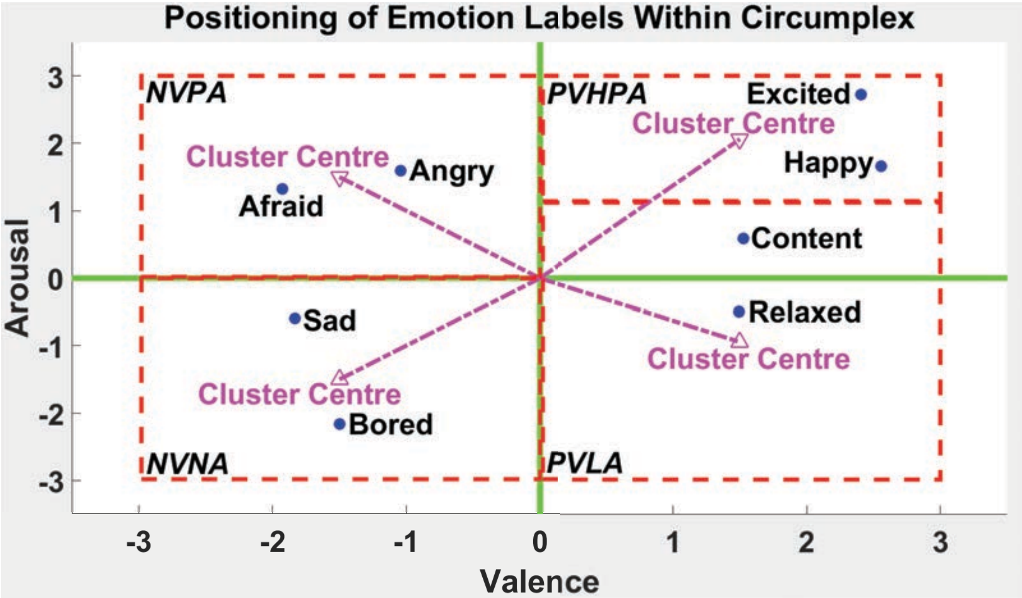


Figure 7 – Subjective Representation of the 8 Verbal Emotions, within the *2D Affective Space* – Clusters are named as follows: (1) Positive Valence Low Arousal *(PVLA)* – (2) Positive Valence High Positive Arousal *(PVHPA)* – (3) Negative Valence Positive Arousal *(NVPA)* – (4) Negative Valence Negative Arousal *(NVNA)*

Table 3 – Affective Clusters Definition, According to the Clusters' Centroid, Lower and Upper Limits

Affective Cluster	Centroid (Cluster's Defined Lower Limit – Cluster's Defined Upper Limit)	
PVLA	Valence	1.5 (0 – 3)
	Arousal	-0.85 (-3 – 1.16)
PVHPA	Valence	1.5 (0 – 3)
	Arousal	2.14 (1.16 – 3)
NVPA	Valence	-1.5 (-3 – 0)
	Arousal	1.5 (0 – 3)
NVNA	Valence	-1.5 (-3 – 0)
	Arousal	-1.5 (-3 – 0)

2.4. Indirect Emotion Labels Positioning

In this study the emotional experiences of 113 participants, exposed to various number of Affective VR environments, have been self-assessed and recorded (Preliminary, Primary and Evaluation Experiments – Chapters 3, 5 and 6). At the end of each Affective VR experience the participants were instructed to report their emotions according to both dimensional (Valence, Arousal and Dominance) and categorical (Relaxed, Content, Happy, Excited, Angry, Afraid, Sad and Bored) models. Although in the self-assessments, the participants reported their emotional experiences within a particular Affective VR experience; their dimensional and categorical ratings could be considered as an *indirect* and unintended positioning of the Emotion Labels within the 3-dimensional affective space (Valence *vs*. Arousal *vs*. Dominance ratings, beside the corresponding equal Emotion Label). This is in contrast with the survey experiment (Experiment 1 – Section 2.3.2), where the participants were asked to (*directly*) position each Emotion Label within the 3-dimensional affective space.

Table 4 – Mean Value of Valence and Arousal Ratings of Emotion Labels and Affective Clusters, across all Participants' Affective VR Ratings – (A - B) Presents the (A) 25th and (B) 75th Percentiles, which Present the Range, and which Contain at Least 50% of the Ratings

Emotion	Mean (25th – 75th Percentiles)		Cluster	Mean (25th – 75th Percentiles)	
Relaxed	Valence	0.98 (0 – 2)	**PVLA**	Valence	1.04 (1 – 2)
	Arousal	-0.98 (-2 – 0)			
Content	Valence	1.07 (1 – 2)		Arousal	-0.01 (-1 – 1)
	Arousal	0.54 (0 – 1)			
Happy	Valence	1.79 (1 – 2)	**PVHPA**	Valence	1.82 (1 – 2)
	Arousal	1.53 (1 – 2)			
Excited	Valence	1.83 (1 – 2)		Arousal	1.93 (1 – 3)
	Arousal	2.24 (2 – 3)			
Angry	Valence	-1.56 (-2 – -1)	**NVPA**	Valence	-1.48 (-2 – -1)
	Arousal	1.26 (1 – 2)			
Afraid	Valence	-0.72 (-2 – 0)		Arousal	1.28 (1 – 2)
	Arousal	1.43 (1 – 2)			
Sad	Valence	-1.29 (-2 – -0.5)	**NVNA**	Valence	-1.17 (-2 – 0)
	Arousal	-0.09 (-2 – 1)			
Bored	Valence	-1.15 (-2 – 0)		Arousal	-1.11 (-2 – 0)
	Arousal	-1.35 (-2 – -1)			

In total, 1827 ratings were recorded within the Indirect Emotion Labels positioning database. We performed an analysis on these ratings to compare the direct and indirect positioning of Emotion Labels, within the dimensional affective space.

According to a MANOVA analysis[13], the dimensional ratings of the indirect positioning of Emotion Labels are significantly different across Emotion Labels ($P_{\text{Labels}} < 0.001$ – similar to direct positioning – Section 2.3.5). On the other hand, an Analysis of Variance (ANOVA) on each axis[14] highlighted that the positioning of Emotion Labels on Valence and Arousal axes, in direct and indirect positioning, were not significantly different ($P(\text{Valence})_{\text{Direct-}vs\text{-indirect}} = 0.062$ and $P(\text{Arousal})_{\text{Direct-}vs\text{-indirect}} = 0.934$), whereas they were significantly different in Dominance axis ($P(\text{Dominance})_{\text{Direct-}vs\text{-indirect}} = 0.01$). This could conclude that the participants followed the Emotion Labels' positioning definitions, presented in the direct positioning, when assessing their emotional experience, in Valence and Arousal axes; whereas they followed a significantly different pattern in Dominance axis. A 3-dimensional MANOVA analysis[15] highlighted that the direct and indirect positioning of Emotion Labels are significantly different ($P_{\text{Direct-}vs\text{-indirect}} = 0.003$) in 3-dimensional Affective Space. However, a 2-dimentional MANOVA analysis[16] highlighted that the direct and indirect positioning of Emotion Labels are **not** significantly different ($P_{\text{Direct-}vs\text{-indirect}} = 0.171$) in 2-dimensional Circumplex of Affect.

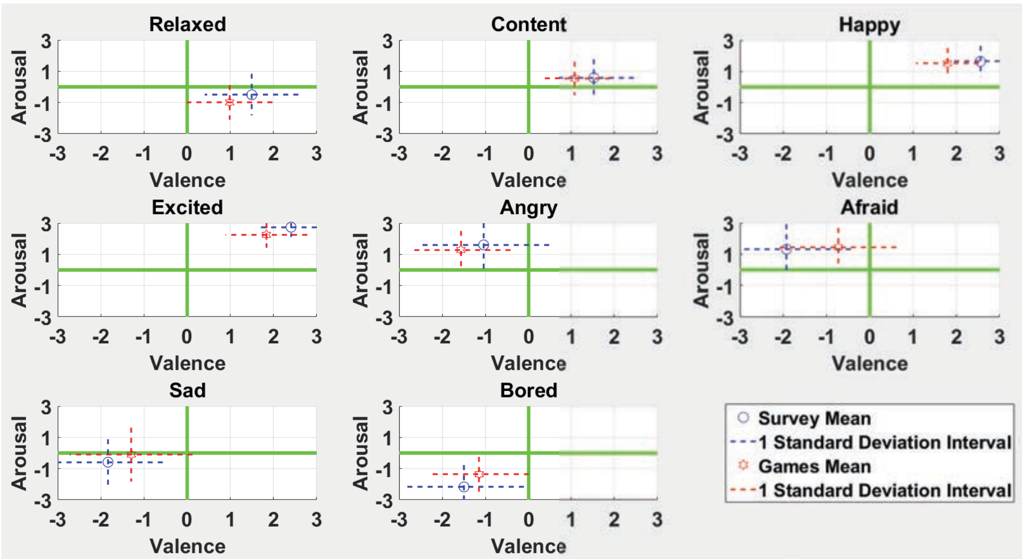


Figure 8 – Direct vs. Indirect Labels Positioning Comparison – The Circles and Hexagrams Present The Labels Mean Positions According to (Respectively) Survey (Direct) and Affective VR (Indirect) Ratings – The Dashed Lines Present the 1 Standard Deviation Interval

As mentioned in Section 2.3.5, we decided to disregard the Dominance axis and employ the 2-dimensioanl Circumplex of Affect. Therefore, as the direct and indirect positioning of Emotion Labels are not significantly different in 2-dimensional Circumplex of Affect, one could conclude that the participants' dimensional and categorical ratings, under emotional experiences, followed the same relationship patterns, highlighted in the direct positioning of Emotion Labels (Section 2.3.5).

[13] Valence and Arousal as dependent variables, while Emotion Labels fixed factor.
[14] One of the Axes ratings (Valence, Arousal or Dominance) as dependent variables, while different positioning (direct *vs*. indirect) as fixed factors.
[15] Valence, Arousal and Dominance as dependent variables, while different positioning (direct *vs*. indirect) as fixed factors.
[16] Valence and Arousal as dependent variables, while different positioning (direct *vs*. indirect) as fixed factors.

Figure 8 presents a comparison between the direct and indirect positioning of the Emotion Labels. Table 4 presents the mean and ranges of the Valence and Arousal ratings of the indirect Emotion Labels and Affective Clusters positions (compare Table 1 and Table 4 for direct *vs*. indirect comparison).

2.5. Conclusion

In this chapter, the definition of emotions and their representation models (categorical *vs*. dimensional) have been discussed in detail (Section 2.1). Moreover, the processes in which the emotional experiences of human could be assessed and labelled (regardless of the employed affective model) have also been considered (Section 2.2). As discussed in Section 2.2, we believe that self-assessment techniques, specifically for the purposes of emotion assessment, can be considered as the most convenient technique in the evaluation process of affective experiments. Furthermore, by conducting a survey, the relationship between the 3D affective space (dimensional model), and eight Emotion Labels (categorical model) has been investigated (Section 2.3). The results highlighted a similar distribution of labels within the dimensional space, when compared to the 2D Circumplex of Affect, presented by Russell (Section 2.3.3). However, not only the survey results, but also the datasets recorded in this thesis (Preliminary, Primary and Evaluation Experiments – Chapter 3, 5 and 6), and other affective databases (IAPS and IADS) presented in the international literatures, showed a consistent high and significant correlation between the Valence and Dominance axes (Section 2.3.4). These results suggested that the 2D Circumplex of Affect needs to be employed in the analysis, as the Dominance axis could not provide any additional affective information (Section 2.3.5). Finally, by analysing the datasets recorded in this thesis (Preliminary, Primary and Evaluation Experiments – Chapter 3, 5 and 6), the indirect and direct distributions of the Emotion Labels, within the 3D affective space, have been compared (Section 2.4). The analysis suggested that the direct subjective allocation of the Emotion Labels within the 2D Circumplex of Affect (Valence *vs*. Arousal) were similar to their indirect allocation (conducted unintentionally by participants, while assessing their emotional experience within the VRs, according to both dimensional and categorical models). This means that the participants have been able to follow almost the same 2D (Valence *vs*. Arousal) distribution pattern for Emotion Labels, presented in their direct survey, when assessing their emotional experiences, within the Affective VRs. This also means that the defined Affective Clusters, according to the 2D distribution pattern of Emotion Labels (Figure 7), are fairly reliable to cluster the emotional experience of the participants, accordingly.

The analyses and findings of this chapter are the foundations of the conceptualisation, designing and evaluation process of the affective computing system (discussed in Chapters 3, 4, 5 and 6). To be able to conduct any psychological or physiological investigation on human emotional behaviour, the foundation, in which the affects are modelled, assessed and represented, need to be explored and clarified first (research question I presented in Section 1.8).

The most important contribution of this chapter is the distribution of Emotion Labels, within the 3D affective space, according to VR-based affective experiences. As suggested by Russell, and Mehrabian, the distribution of emotion labels, within either the 2D or 3D model, would differ between people with different cultures, especially those with different languages (Russell, 1980; Mehrabian, 1970). Although the labels followed the same order, presented within the Circumplex of Affect, it was highlighted that ‘Relaxed’ and ‘Content’ are associated with higher arousal levels (compared to the Circumplex of Affect). It also underlined the fact that the Circumplex of Affect, according to VR-based affective experiences, cannot be clustered into 4 quadrants, as 4 Affective Clusters (one of which sharing two quadrants) are required to categorise the continuous affective space.

Another contribution of this chapter is the identification of the correlation between affective axes. Mehrabian and Russell employed a multi-question subjective questionnaire to model the emotional behaviour of people. By employing principal component analysis (PCA), Mehrabian extracted three components related to the level of Valence, Arousal and Dominance, to represent the affective space. Whereas Russell concluded that employing Valence and Arousal levels are sufficient, to represent human emotions (Russell, 1980; Mehrabian, 1970). In this chapter, we encountered a significant high correlation between Valence and Dominance axes, in 4 different independent affective datasets (two recorded in this study and two presented in international literatures). From this one can conclude that Russell’s 2D Circumplex of Affect can represent the affective space more efficiently than Mehrabian’s 3D affective space.

Chapter 3
Affective Virtual Reality

Abstract – Detecting and measuring emotional responses whilst interacting with Virtual Reality (VR), and assessing and interpreting their impacts on human engagement and "immersion", are both academically and technologically challenging. Whilst many researchers have, in the past, focused on the affective evaluation of passive environments, such as listening to music or the observation of videos and imagery, virtual realities and related interactive environments have only been used in a small number of research studies as a mean of presenting emotional stimuli. This chapter reports the first stage (focusing on participants' subjective responses) of a range of experimental investigations supporting the evaluation of emotional responses. To populate the affective space with participants' emotional responses, an "Affective VR", capable of manipulating users' emotions, has been designed and subjectively evaluated. The VR takes the form of a dynamic "speedboat" simulation, elements (controllable VR parameters) of which were assessed and selected based on a 35-respondent online survey, coupled with the implementation of an affective power approximation algorithm. A further 68 participants took part in a series of trials, interacting with a number of VR variations, while subjectively rating their emotional responses.

3. Affective Virtual Reality

To perform the affective recognition process in virtual realities, a system has to be designed, trained and validated with respect to a psychophysiological affective database, recorded from a large number of users, exposed to a number of controlled and known affective stimuli (considering supervised learning algorithms (Mohri et al., 2012)). To construct such a database, a number of controlled emotional situations, evoking some specific affective states on the part of the users[17], would 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, would then be analysed for the design, training and validation of the affective recognition system. Therefore, two distinct steps in the psychophysiological affective database construction can be considered: (a) evoking controlled emotional experiences and (b) the measurement of physiological parameters. It would be important to ensure the implementation of strict experimental designs in such a paradigm, in order to avoid the development of an inappropriate psychophysiological affective database, which would invalidate the recognition system's training process. As an illustration, if the users' emotional experiences were poorly controlled (e.g., it was not possible to state with confidence that anger had been experienced by the users during the corresponding session), then the classification techniques would be unable to train the affective recognition system properly and the accuracy of the system would be affected accordingly. To prevent such incidents, the emotional stimuli must be subjectively evaluated and categorised prior to the undertaking of physiological measurements, in order to validate their effectiveness in evoking the required emotional experiences on the part of all users.

To date a number of evaluated affective stimuli databases using images (the International Affective Picture System – IAPS (Lang et al., 2008)), sounds (the International Affective Digital Sounds – IADS (Bradley & Lang, 1999)) and video clips (Baveye et al., 2013) have been presented in the literature. These established databases provide investigators with a variety of pre-evaluated affective stimuli, which (from a subjective outcome perspective) have been found to elicit specific (and quite strong) emotions in recipients. However, to the knowledge of the authors, no validated Affective VR-based stimuli database has been presented as yet. The availability and reliability of such a database of stimuli in form of a virtual reality (VR), is crucial in the design and validation of an affective computing system, which can be used in VR-based systems. In the present chapter, an *Affective Virtual Reality* and the process by which it was conceptualized, designed and subjected to an early validation study is discussed in detail. The Affective VR is capable of eliciting multiple emotions within the users, and manipulating their affective experiences within the affective space, by controlling the VR's internal parameters. A number of "sub-games" (based on the selection of multiple unique VR parameters) have been

[17] Such as images that evoke fear and disgust on users – image number 3000 to 3266 in IAPS (Lang et al., 2008)

selected using an affective power estimation process, and evaluated in a subjective experiment employing 68 participants.

3.1. Affective VR Design

Two different approaches are possible for designing an Affective Virtual Reality, capable of evoking certain emotions:

1. **Multiple VRs:** A number of **entirely different** VRs can be designed to evoke different emotions. The first advantage of this approach is that every VR can be designed in a way that would have the maximum impact on the users, in order to evoke a particular emotion. The disadvantage of this method is the variability between the environments. Different environments may well result in different VR experiences, which may in turn lead to too much variability between the recorded data. This would leave no ground truth for any comparison between the independent situations. Also, each VR would take the form of a new environment for the participants and could create an element of surprise in every attempt. This issue may decrease or even change the expected emotional experience on the part of the participants.
2. **Single VR:** A single but well constructed virtual reality can be designed that is capable of evoking different emotions by changing the simulated environment's internal properties. The first advantage of this approach is the minimum variability between the emotional experiences, as the background environment or scenario (or called in the thesis *Neutral Scenario* of *VR* – the overall theme, environment, interaction process and rules of the VR) for all experiences would be the same, and changes in the parameters of the VR and incidents could elicit different emotions. Another advantage would be the minimum element of surprise on the participants (compared to the multiple-VRs approach). The overall VR environment, interaction algorithm and other aspects related to the background scenario would stay the same and allow participants to concentrate on the affective parameters rather than the changes. The disadvantage of this approach is that the effectiveness of emotional experiences may be less influential than the first method. The reason for this is that, in a multiple VR approach, one scenario can be designed to evoke boredom and another to elicit excitement; each in a very powerful way, while in this approach there is only one VR scenario, which should be capable of evoking all emotions in an effective manner.

In human-centred experimentation, minimum variability within a single VR experience is an extremely important matter, as any acceptable analysis dealing with either affects or physiological databases has to be based on changes in emotional experiences, due to *different environments* (between sub-games), rather than different *personal experiences* (between participants). Multiple VRs may create different experiences amongst participants, rather than a single VR with an overriding context, due to variability between environments. Therefore, in the present project, the **Single**

VR approach has been adopted as the design approach for the virtual affective medium.

A *Speedboat Simulation*[18] game (Figure 9) was designed for use as the background scenario of the Affective VR. As the experimental cohort was anticipated to comprise both gamers and non-gamers, it was decided to use a driving-based simulation with a simple directional interface style (a speedboat scenario in this case), to reduce the amount of prior gaming experiences required for participation (i.e. when compared to the skills and experiences typically manifested by players of first person combat and strategic games). Also, the creation of an environment with a very basic contextual setting in terms of graphical elements drove the choice of a speedboat simulation (as opposed to automobile driving, which typically consists of complex urban representations). Moreover, the dynamics of the environment would, it was felt, provide a wide range of possible parameters and variables that could be implemented and controlled in the environment (described in more detail in Section 3.2).

Figure 9 – Speedboat Simulation Environment

In the *neutral* speedboat simulation environment, the user is able to navigate a small boat, freely, within a coastal virtual environment, originally created for VR healthcare research (Stone & Hannigan, 2014). By manipulating the VR parameters (described in more detail in Section 3.2), a number of different variations of the neutral environment were created. These variations have been called *sub-games*, in this study. In majority of sub-games, there are a number of floating "ring buoys" on the water that the users can collect to gain higher scores. In all sub-games, the users can either finish the game by passing the finish line at a distance (from where the game is started), or continue exploring for as long as they require (only in the sub-games, which do not have any time limitations). Regardless of the time settings for each VR variation, no sub-game is allowed to continue beyond 5 minutes. If the participant spends longer than 5 minutes in a particular sub-game, it terminates automatically. Depending on the VR settings (Section 3.2), participants can interact with the virtual environment using either a mouse or a force feedback joystick. The

[18] This simulation can be viewed at: https://www.youtube.com/watch?v=pqn-X1Z5AoM

joystick is capable of displaying vibration effects according to simulated “water turbulence” and, in addition, left/right forces on the grip, simulating simple “water resistance” effects, created when the boat is turning. A Samsung 32-inch flat LCD screen was used to present the VR scenes, together with a Sennheiser RS-170 wireless headphone to play the environmental sound effects.

3.2. Affective Parameters (“Incidents”)

In order to evoke multiple emotions in the participants, a number of controllable affective parameters need to be identified and implemented within the VR. Manipulation of these parameters would (it was hypothesized) evoke different emotions within the participants. The general nature of these parameters or “incidents” needs to be studied prior to any identification or implementation within the environment.

3.2.1. Categorization of Incidents

I. VR Aspects: For the purposes of this study, the parameters or incidents presented in the speedboat simulation were categorized into 4 major aspects:

1. **Visualization:** Any aspect of the game related to visual stimulation, including lighting, textures, fidelity, scale of the objects, realism of any action (such as avatar animation) and physical behaviours.
2. **Auditory:** All features of the game that are related to the auditory sense of the users, including the background music, sounds of objects, voices of avatars, and so on.
3. **Interaction:** Keyboard, mouse, joysticks, voice recognition systems, gestural translators, and so on, all fall within the interaction category.
4. **Narrative:** Any aspect of the game (visual, auditory and interaction) that is presented to the users in a meaningful or contextually relevant way through a narrative or background scenario. This aspect can influence the user’s perception and change his or her experience quite dramatically. As an illustration, even a game created using extraordinary visualizations, auditory and interaction factors can have different influences on users simply by the way the game’s narrative has been presented. If the background presents a science fiction scenario, for example, the user may expect to experience extreme levels of action, a high tempo, even fear. Yet, the same game presented with a real-life scenario as its background narrative, perhaps one that depicts a desert island or peaceful countryside setting, can create a completely different set of expectations and perceptions on the part of the user.

II. Timing Aspects: Each incident in the game can be presented to the users either as a single event in the game (“In-Game Discrete Event”, such as a sudden sound, a short aggressive attack, a short screen vibration, etc.), or throughout the whole duration of a game (“Game-Persistence”, such as a time limitation, a change in the input device control law, etc.).

Considering the classification of the in-game parameters, according to the VR and time related aspects, the in-game parameters, which can be game-persistence, could be considered as “Parameters”. However, the in-game discrete events have to be considered as “Incidents”, as their short-time presence within a sub-game cannot classify them as an in-game parameter.

3.2.2. Incidents Identification, and Assessment

According to the speedboat simulation’s environmental capabilities, 21 possible incidents were identified for implementation within the Affective VR. These incidents were categorised based on their presentation timing, together with the VR aspects. Table 5 shows these incidents, clustered with respect to the VR aspect and timing classifications. Different combinations of these incidents could create different sub-games. Combination of elements within the columns can create 1444 different sub-game combinations (C_i means i^{th} column $-\ 5(C_1)\times2(C_2)\times2(C_3)\times2(C_4)\times3(C_5)\times2(C_6)\times2(C_7)\times3(C_8) = 1440$). As some of these combinations are not possible (e.g., no time limitation while the timer is faulty, etc.), the total number of possible combinations is, as a consequence, reduced to 792 different sub-games.

Table 5 – 21 Incidents Categorization According to VR and Timing Aspect

Narrative Based				Interaction Based		Visualisation Based	
Sub-Scenario	Time	Timer Functionality	Barrier	Controller Type	Controller Functionality	Camera Movement	Screen Color
Free Environment Exploration	Time Limitation	Normal Timer	No Invisible Barrier	Mouse	Normal Controller	Shaking and Blurring the Camera	Color Screen
Mine Avoidance	No Time Limitation	Faulty Timer	Invisible Barrier	Joystick Without Force Feedback	Faulty Controller	No Camera Shake or Blurring	Black and White Screen
Torpedo Avoidance				Joystick With Force Feedback			Inverse Black and White Screen
Shooting Flying Ball							
Maze							
Game-Persistent	Game-Persistent	In-Game Discrete Event	In-Game Discrete Event	Game-Persistent	In-Game Discrete Event	In-Game Discrete Event	Game-Persistent

3.3. Affective Virtual Reality

The speedboat VR is capable of generating all required combinations of incidents (described in Section 3.2.2). Figure 10 presents some examples of possible sub-game combinations. Each sub-game was allocated an 8-digit code. Each code represented the index number within each column of Table 5. As an illustration, code "21223111" would set up a sub-game environment with the following settings:

Mine Avoidance + Time Limitation + Faulty Timer + Invisible Barrier + Joystick with Force Feedback + Normal Controller + Shaking and Blurring the Camera + Colour Screen

Figure 10 – Incidents Presentation Examples Within the Affective VR – A) Start Line Flag for Time Limitation Scenarios – B) Ramps in the Virtual Environment – C) Mine Avoidance – D) Jumping Over Ships – E) Deriving Freely Outside the Ring Buoys Lane – F) Finding Hidden Ring Buoys Inside the Bushes, By Using the Radar – G) Torpedo Avoidance – H) Splashing Water to the Flying Ball – I) Inverse Black and White Screen in the Torpedo Avoidance Sub-Scenario – J) Black and White Screen in Maze Sub-Scenario – K) Finish Line Flag to Terminate the Game on Demand – L) Score Calculation at the Game Termination

The experimenter generates a list of these codes for the VR, in order to create an automated sequential (randomised) experiment. In addition, an interactive questionnaire for assessing the affective experience of the participants, according to Valence, Arousal and Dominance axes (scaled between -3 to +3 in all axes), followed by the 8 Emotion Labels list described earlier, were automatically presented to the participants, at the end of each sub-game (Section 2.3.1). The rating results, followed by the sub-game information, were saved in a text file during the run-time of the experiment, and could be simply extracted after the experiment.

3.4. Pre-Experiment Survey (*Experiment 2*)

3.4.1. Sub-Games Affective Power Approximation

As explained above (Section 3.2.2), 792 different sub-games can be constructed using the 21 *incidents*. Two different approaches were available to test the emotional effect of each sub-game:

1. **Subjective Assessment:** In this approach all sub-games need to be played at least once, by one of the participants. It is impossible to allow each participant to play all 792 sub-games, as no one individual would be able to play all of them without experiencing extreme fatigue (even in multiple sessions, over different days). Therefore all 792 sub-games can be divided into “m” sub-sets, each of which contains a number of sub-games. Then each sub-set can be played (I) either by “n” participants, or (II) by only one participant. To be able to perform a meaningful affective analysis, the affective power of sub-games cannot be assessed by subjective assessment of a single participant (II). Therefore, a high number of participants need to be recruited, to enable each game to be played by “n” participants (I).
2. **Subjective Estimation:** In this method, the emotional effect of each incident, rather than each sub-game, would be evaluated, using an approximation technique. This means that each participant would estimate the possible emotional effect of each *incident* (all 21 *incidents* considered), described verbally (Section 3.4.2). Then, by employing the approximated emotional effects of all *incidents,* and an estimation technique (Section 3.4.3), the overall affective power of each single sub-game, containing a number of incidents, can be approximated.

Due to the high number of possible sub-game combinations, to reduce this number to include those affective combinations, which would be most likely to manipulate the participants’ emotional status towards all four Affective Clusters (described in Section 2.3.5), the **Subjective Estimation** approach was employed in this study.

3.4.2. Participants and Method

Subsequently, a subjective survey was designed and presented online to 35 respondents (with a mean age of 24.72 years, and a distribution of 57% males and 66% non-gamers). The study was reviewed and approved by the University of

Birmingham's Ethical Review Committee (Ethical Reference Number: ERN_13-1157). To distinguish gamers from non-gamers, the following description was presented to the respondents as part of the online survey to enable them to self-assess appropriately:

"If you follow games in the market regularly and have a lot of experience playing games on PC and consoles, you are a gamer."

Within the survey, a brief overarching explanation of the VR followed by a short video of the environment (as referenced earlier under Footnote 18) was presented to the respondents. Then, each incident was described in text form, such as: "imagine that you need to drive the boat through mines scattered on the water", or "imagine that the controller used to control the boat is faulty and is not responding to your actions". The respondents were then required to estimate their Valence, Arousal and Dominance levels, and to choose one of the eight Emotion Labels (as presented in Section 2.3.1), for each incident, by considering themselves within the described affective situation (refer to Appendix C).

3.4.3. Results and Discussion

Using the mean ratings (across participants) for each incident, the affective powers of all VR parameters have been approximated within the 3-dimensional affective space and are shown in Figure 11.

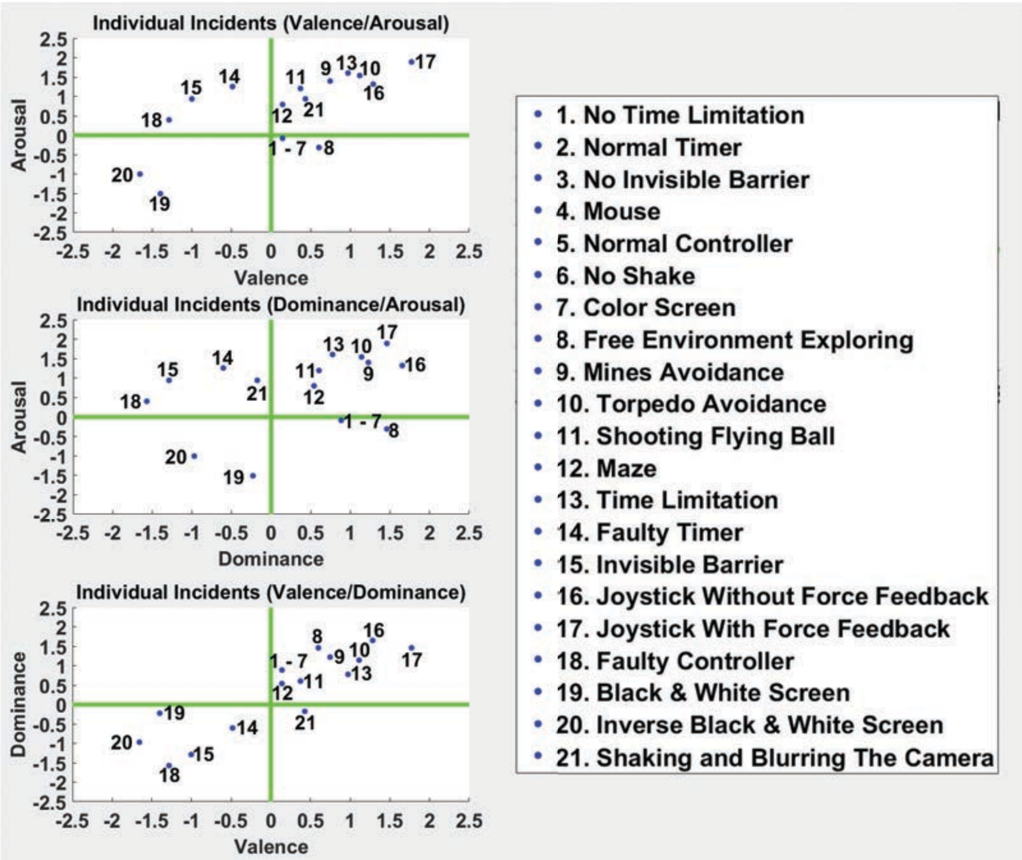


Figure 11 – Presentation of Estimated Affective Power of the Incidents Within the 3D Affective Space

To analyse and estimate the total emotional power of each **sub-game**, an estimation algorithm was designed based on 2 hypotheses:

1. **Interacting and Additive Effect:** Each individual incident can have an effect on another incident if they are both presented within the same sub-game. This means that incidents can have additive effects on each other. It also means that, if several incidents are presented in a sub-game, the overall emotional effect of that combination can be considered as the summation of Circumplex values of all individual incidents.
2. **Background Game as the Neutral:** The background scenario can be considered as neutral, with (0, 0, 0) as its 3D emotional effect. This means that all possible combinations would be evaluated with respect to the background VR scenario.

Accepting these hypotheses then, the affective power of all 792 sub-games can be estimated by adding the approximated 3D affective values of all incidents within each combination (sub-game). Figure 12 presents the positioning of all 792 sub-games within the 3-dimensional affective space. Furthermore, to estimate the "*Occurrence Probability*" (*OP*) of each categorical label for each game in the future subjective experiment, Equation 1 was employed in the analysis. By using this equation the probability, in which a particular label can be selected in a specific sub-game is approximated.

$$OP = \frac{\sum(Emotion\ Label\ Occurence\ Frequency\ of\ all\ incidents\ in\ a\ subgame)}{(Number\ of\ Incidences\ within\ the\ subgame) \times (Number\ of\ participants\ in\ the\ experiment)}$$

Equation 1 – Categorical Label Occurrence Probability Estimation Formula

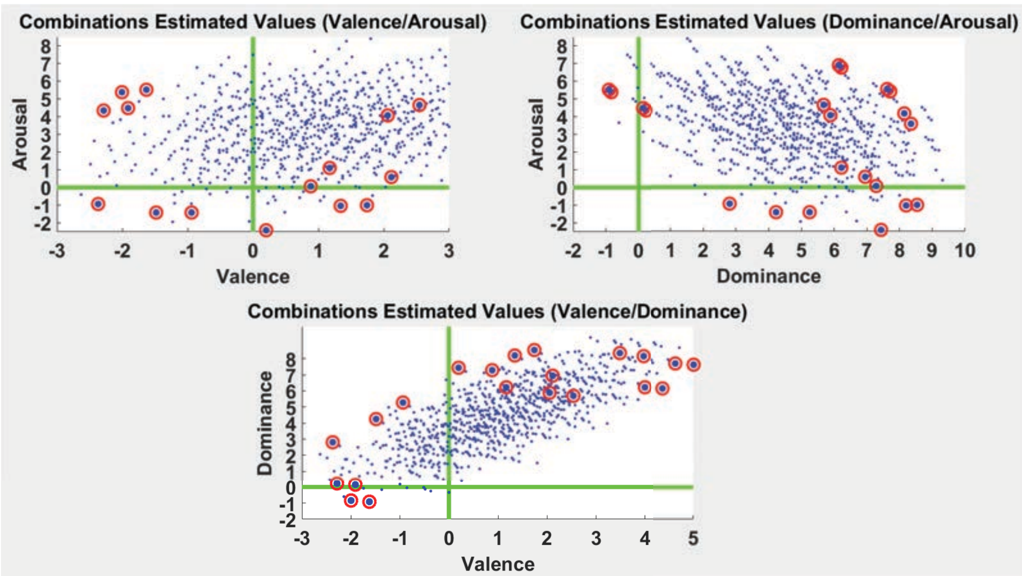


Figure 12 – Presentation of Sub-Games Within the 3D Affective Space – Dots Represent the sub-games – Circled Dots Represents the 22 Selected Sub-Games

3.5. Preliminary Subjective Experiment (*Experiment 3*)

3.5.1. Sub-Games Selection for Preliminary Experiment

After performing the affective power estimation process on the sub-games, the post-exposure subjective affective response of the participants to a number of sub-games can be assessed. This subjective affective evaluation can, firstly, assess the accuracy of the approximation process and secondly, identify the subjective emotional power of a number of sub-games (rather than their estimated values), for future affective experiments. Therefore, a number of sub-games need to be selected to be presented to a number of participants, for subjective affective evaluation.

For this experiment, it was decided to adopt an overall "*Experiment Duration*" (*ED*) of less than 2 hours (in order to minimise participant fatigue). Considering the maximum duration of each sub-game as 5 minutes (no sub-game was permitted to take longer than 5 minutes, Section 3.1, although they usually lasted less than this), and the training session (Section 3.5.3) duration of between 5 and 15 minutes (average 10 minutes), the maximum number of sub-games for the experiment (not to exceed 2 hours experiment duration), was calculated as 22 (Equation 2). Therefore the 22 *Most Affective Sub-Games* (those, which are most likely to evoke emotional experience within the four affective clusters – this term is used throughout the thesis) have to been identified (using their approximated values) to be presented in the experiment.

$$ED = n \times 5(\text{minutes}) + 10(\text{minutes}) \leq 2(\text{hours})$$
$$n = number\ of\ subgames$$
$$n \leq 22$$

Equation 2 – Calculation of the Maximum Required Number of Sub-Games

3.5.2. Most Affective Sub-Games Selection

In order to select the most affective combinations, each sub-game was presented as a vector, in a 7-dimensional space, presented in Table 6. The Valence, Arousal and Dominance values were calculated through the sub-game affective power estimation algorithm (Section 3.4.3). In addition, the approximated occurrence probability (*OP*) for each sub-game was used to create the sub-games' affective vectors (Section 3.4.3). In an ideal situation, the most affective sub-game within each cluster would feature that cluster's central values for Valence, Arousal and Dominance (ideally, in the dimensional model – Section 2.3.3 – the clusters' *centroids*); while all participants have chosen one of the two verbal labels within that cluster (i.e. the probability of selecting either of the cluster's labels is 100% – ideally in the categorical model). Therefore the clusters' ideal vectors could be presented as shown in Table 7.

Table 6 – 7-Dimensional Presentation of Sub-Games' Vectors

Valence	**Arousal**	**Dominance**	**PVLAPD Occurrence Percentage**	**PVHPAPD Occurrence Percentage**	**NVPAND Occurrence Percentage**	**NVNAND Occurrence Percentage**
Valence Mean Value Across All Participants	Arousal Mean Value Across All Participants	Dominance Mean Value Across All Participants	Fraction of Participants Who Chose Either Relaxed or Content	Fraction of Participants Who Chose Either Happy or Excited	Fraction of Participants Who Chose Either Angry or Afraid	Fraction of Participants Who Chose Either Sad or Bored

Table 7 – 7-Dimensional Presentation of Clusters' Ideal Vectors

Cluster	**Valence**	**Arousal**	**Dominance**	**PVLAPD Occurrence Percentage**	**PVHPAPD Occurrence Percentage**	**NVPAND Occurrence Percentage**	**NVNAND Occurrence Percentage**
PVLAPD	1.5	-0.85	1.5	100%	0%	0%	0%
PVHPAPD	1.5	2.14	1.5	0%	100%	0%	0%
NVPAND	-1.5	1.5	-1.5	0%	0%	100%	0%
NVNAND	-1.5	-1.5	-1.5	0%	0%	0%	100%

However, as it was discussed in Section 2.3.5, it was decided to discard the Dominance axes in the analysis. Therefore the 7-dimensional space, presented above, has been reduced to a 6-dimensional space, by removing the Dominance level. To select the most affective sub-games, the Cosine Similarity algorithm (Pang-Ning Tan, 2005) was employed to find the 4 most similar sub-game affective vectors to the clusters' ideal vectors in each Affective Cluster (refer to Appendix D). Therefore, in each cluster the 4 most powerful combinations were selected to consider the 16 most affective sub-games, which can cover the entire affects space effectively. Furthermore, 5 additional test combinations (added manually – those, which have the maximum standard deviation and minimum level of agreement among participants), followed by the *neutral sub-game* (The sub-game with background scenario settings – "12111121" combination) were added to create the 22-game experiment. Figure 12 presents the 22 selected sub-games among 792 combinations, highlighted with circles. Table 8 presents the arrangement of the incidents within the 22 selected sub-games.

Table 8 – The 22 Selected Sub-Games' Settings

	Narrative					Interactive		Visualization	
#	Game Code	Main Scenario	Time Limitation	Timer	Invisible Barrier	Controller Type	Faulty Controller	Camera	Screen Colour
1	12111121	Free Environment Exploring	No Time Limitation	Normal Timer	No Invisible Barrier	Mouse	Normal Controller	No Shake or Blurring	Colour Screen
2	12111122	Free Environment Exploring	No Time Limitation	Normal Timer	No Invisible Barrier	Mouse	Normal Controller	No Shake or Blurring	Black & white
3	12111222	Free Environment Exploring	No Time Limitation	Normal Timer	No Invisible Barrier	Mouse	Faulty Controller	No Shake or Blurring	Black & white
4	12111223	Free Environment Exploring	No Time Limitation	Normal Timer	No Invisible Barrier	Mouse	Faulty Controller	No Shake or Blurring	Inverse Black & White
5	12112122	Free Environment Exploring	No Time Limitation	Normal Timer	No Invisible Barrier	Joystick Without Force Feedback	Normal Controller	No Shake or Blurring	Black & white
6	12113122	Free Environment Exploring	No Time Limitation	Normal Timer	No Invisible Barrier	Joystick With Force Feedback	Normal Controller	No Shake or Blurring	Black & white
7	12121122	Free Environment Exploring	No Time Limitation	Normal Timer	Invisible Barrier	Mouse	Normal Controller	No Shake or Blurring	Black & white
8	12121222	Free Environment Exploring	No Time Limitation	Normal Timer	Invisible Barrier	Mouse	Faulty Controller	No Shake or Blurring	Black & white
9	21113111	Mine Avoidance	Time Limitation	Normal Timer	No Invisible Barrier	Joystick With Force Feedback	Normal Controller	Shaking and Blurring the Camera	Colour Screen
10	21213111	Mine Avoidance	Time Limitation	Faulty Timer	No Invisible Barrier	Joystick With Force Feedback	Normal Controller	Shaking and Blurring the Camera	Colour Screen
11	21221223	Mine Avoidance	Time Limitation	Faulty Timer	Invisible Barrier	Mouse	Faulty Controller	No Shake or Blurring	Inverse Black & White

12	21221213	Mine Avoidance	Time Limitation	Faulty Timer	Invisible Barrier	Mouse	Faulty Controller	Shaking and Blurring the Camera	Inverse Black & White
13	31113111	Torpedo Avoidance	Time Limitation	Normal Timer	No Invisible Barrier	Joystick With Force Feedback	Normal Controller	Shaking and Blurring the Camera	Colour Screen
14	31213111	Torpedo Avoidance	Time Limitation	Faulty Timer	No Invisible Barrier	Joystick With Force Feedback	Normal Controller	Shaking and Blurring the Camera	Colour Screen
15	31221223	Torpedo Avoidance	Time Limitation	Faulty Timer	Invisible Barrier	Mouse	Faulty Controller	No Shake or Blurring	Inverse Black & White
16	31221213	Torpedo Avoidance	Time Limitation	Faulty Timer	Invisible Barrier	Mouse	Faulty Controller	Shaking and Blurring the Camera	Inverse Black & White
17	41112121	Shooting a Flying Ball	Time Limitation	Normal Timer	No Invisible Barrier	Joystick Without Force Feedback	Normal Controller	No Shake or Blurring	Colour Screen
18	41112221	Shooting a Flying Ball	Time Limitation	Normal Timer	No Invisible Barrier	Joystick Without Force Feedback	Faulty Controller	No Shake or Blurring	Colour Screen
19	41113121	Shooting a Flying Ball	Time Limitation	Normal Timer	No Invisible Barrier	Joystick With Force Feedback	Normal Controller	No Shake or Blurring	Colour Screen
20	41113221	Shooting a Flying Ball	Time Limitation	Normal Timer	No Invisible Barrier	Joystick With Force Feedback	Faulty Controller	No Shake or Blurring	Colour Screen
21	52112122	Maze	No Time Limitation	Normal Timer	No Invisible Barrier	Joystick Without Force Feedback	Normal Controller	No Shake or Blurring	Black & white
22	52112112	Maze	No Time Limitation	Normal Timer	No Invisible Barrier	Joystick Without Force Feedback	Normal Controller	Shaking and Blurring the Camera	Black & white

3.5.3. Participants and Method

An experiment was conducted in which the 22 selected sub-games were presented to 68 participants (with a mean age of 24.12 years). The participants consisted of 4 different groups; male gamers, female gamers, male non-gamers and female non-gamers (17 participants for each group – the gaming experience was subjectively assessed by the participants, according to the description presented at Section 3.4.2). The study was reviewed and approved by the University of Birmingham Ethical Review Committee (Ethical Reference Number: ERN_13-1157). Each experiment commenced with a training session (Figure 13), to prepare the participants for every possible incident within the games. The training introduced the game environment to the participants and served to reduce any element of surprise in the games. The sessions were performed in a quiet room. All participants were provided with a 32-inch Samsung LCD display, a Microsoft Wireless Mouse 5000, a Logitech Wingman 3D force feedback joystick and a Sennheiser RS-170 wireless headphone. As the purpose of this experiment was to evaluate the emotional effect and intensity of the variations of the designed Affective VR, to be employed as the affective stimuli in the psychophysiological database construction process (Section 5.2), it was decided to not to use any form of VR headset, as the participants would not be able to wear both EEG sensors and a VR headset at the same time. On average, participants spent 58 minutes playing the games, and 1 hour, 46 minutes to complete the entire experiment. Therefore, on average, participants spent 48 minutes of the experiment to complete the questionnaire, or to rest between the sub-game sessions (refer to Appendix E and F for the "Consent Form" and "Information Sheet").

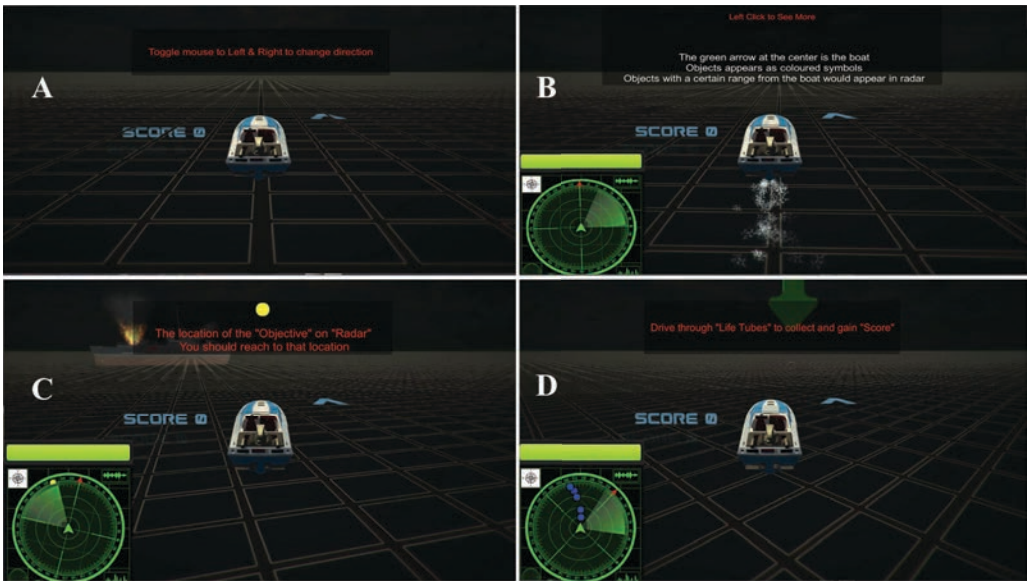


Figure 13 – Affective VR Training Session – A) Practicing the Manoeuvring Procedures – B) Explaining the Radar – C) Describing the Dot Colours in Radar and Their Definitions – D) Presenting the "Ring Buoys" and the Scoring Procedure

Table 9 – Estimated Versus Subjectively Reported Dimensional Levels for the 22 Selected Sub-Games (*SE* is the Standard Error)

	Valence		**Arousal**		**Dominance**	
#	**Estimated**	**Subjective Rate**	**Estimated**	**Subjective Rate**	**Estimated**	**Subjective Rate**
1	1.74 (*SE*=0.26)	1.53 (*SE*=0.12)	-1 (*SE*=0.26)	0.04 (*SE*=0.19)	8.54 (*SE*=0.22)	2.09 (*SE*=0.12)
2	0.2 (*SE*=0.26)	1.23 (*SE*=0.15)	-2.43 (*SE*=0.26)	-0.26 (*SE*=0.21)	7.43 (*SE*=0.23)	2.64 (*SE*=0.09)
3	-2.37 (*SE*=0.26)	-0.57 (*SE*=0.18)	-0.91 (*SE*=0.26)	-0.01 (*SE*=0.19)	2.8 (*SE*=0.24)	-0.39 (*SE*=0.19)
4	-1.49 (*SE*=0.26)	-0.6 (*SE*=0.16)	-1.43 (*SE*=0.26)	-0.15 (*SE*=0.19)	4.23 (*SE*=0.23)	-0.54 (*SE*=0.2)
5	1.34 (*SE*=0.26)	0.91 (*SE*=0.17)	-1.03 (*SE*=0.26)	-0.24 (*SE*=0.22)	8.2 (*SE*=0.22)	2.36 (*SE*=0.12)
6	2.11 (*SE*=0.25)	1.1 (*SE*=0.16)	0.57 (*SE*=0.25)	0.76 (*SE*=0.2)	6.94 (*SE*=0.23)	1.61 (*SE*=0.16)
7	-0.94 (*SE*=0.26)	0.23 (*SE*=0.18)	-1.4 (*SE*=0.26)	0.23 (*SE*=0.19)	5.26 (*SE*=0.23)	1.79 (*SE*=0.18)
8	-2.37 (*SE*=0.26)	-0.76 (*SE*=0.18)	-0.91 (*SE*=0.26)	0.22 (*SE*=0.19)	2.8 (*SE*=0.24)	-0.67 (*SE*=0.2)
9	4.63 (*SE*=0.24)	1.43 (*SE*=0.16)	5.4 (*SE*=0.24)	1.77 (*SE*=0.11)	7.71 (*SE*=0.23)	1.42 (*SE*=0.16)
10	4 (*SE*=0.25)	0.61 (*SE*=0.2)	6.74 (*SE*=0.25)	1.47 (*SE*=0.14)	6.23 (*SE*=0.24)	0.7 (*SE*=0.21)
11	-2.29 (*SE*=0.26)	-1.15 (*SE*=0.18)	4.34 (*SE*=0.26)	0.6 (*SE*=0.2)	0.23 (*SE*=0.25)	-1.12 (*SE*=0.2)
12	-2 (*SE*=0.27)	-1.07 (*SE*=0.18)	5.37 (*SE*=0.27)	0.88 (*SE*=0.17)	-0.83 (*SE*=0.25)	-0.96 (*SE*=0.18)
13	5 (*SE*=0.24)	1.85 (*SE*=0.15)	5.54 (*SE*=0.24)	2.31 (*SE*=0.1)	7.63 (*SE*=0.23)	1.35 (*SE*=0.17)
14	4.37 (*SE*=0.25)	0.8 (*SE*=0.19)	6.89 (*SE*=0.25)	1.76 (*SE*=0.14)	6.14 (*SE*=0.24)	1.09 (*SE*=0.18)
15	-1.91 (*SE*=0.26)	-1.28 (*SE*=0.2)	4.49 (*SE*=0.26)	1.03 (*SE*=0.16)	0.14 (*SE*=0.25)	-1.55 (*SE*=0.19)
16	-1.63 (*SE*=0.27)	-1.14 (*SE*=0.19)	5.51 (*SE*=0.27)	1.39 (*SE*=0.16)	-0.91 (*SE*=0.26)	-1.53 (*SE*=0.17)
17	3.49 (*SE*=0.25)	1.31 (*SE*=0.15)	3.6 (*SE*=0.25)	1.24 (*SE*=0.14)	8.34 (*SE*=0.22)	1.81 (*SE*=0.16)
18	2.06 (*SE*=0.25)	-0.45 (*SE*=0.21)	4.09 (*SE*=0.25)	1.08 (*SE*=0.19)	5.89 (*SE*=0.23)	-1.14 (*SE*=0.19)
19	3.97 (*SE*=0.24)	1.54 (*SE*=0.16)	4.17 (*SE*=0.24)	1.69 (*SE*=0.13)	8.14 (*SE*=0.23)	1.53 (*SE*=0.15)
20	2.54 (*SE*=0.24)	-0.18 (*SE*=0.23)	4.66 (*SE*=0.24)	1.39 (*SE*=0.15)	5.69 (*SE*=0.23)	-1.04 (*SE*=0.18)
21	0.89 (*SE*=0.26)	-0.53 (*SE*=0.23)	0.09 (*SE*=0.26)	-0.33 (*SE*=0.23)	7.29 (*SE*=0.23)	1.7 (*SE*=0.17)
22	1.17 (*SE*=0.26)	-1.11 (*SE*=0.22)	1.11 (*SE*=0.26)	-0.48 (*SE*=0.23)	6.23 (*SE*=0.23)	1.27 (*SE*=0.22)

Table 10 – Estimated Versus Subjectively Reported Categorical Levels for the 22 Selected Sub-Games

#		Relaxed	Content	Happy	Excited	Angry	Afraid	Sad	Bored
1	Estimated Percentage	18.41%	14.60%	10.79%	17.46%	5.08%	2.54%	0.00%	31.11%
	Reported Percentage	23.53%	36.76%	17.65%	14.71%	0.00%	1.47%	0.00%	5.88%
2	Estimated Percentage	17.14%	13.65%	9.52%	15.56%	5.71%	2.22%	2.22%	33.97%
	Reported Percentage	28.79%	21.21%	28.79%	1.52%	3.03%	0.00%	0.00%	16.67%
3	Estimated Percentage	15.87%	12.70%	8.25%	13.97%	13.01%	2.22%	2.54%	31.43%
	Reported Percentage	7.46%	13.43%	10.45%	5.97%	28.36%	1.49%	4.48%	28.36%
4	Estimated Percentage	15.24%	12.38%	9.21%	13.97%	14.92%	3.49%	1.90%	28.89%
	Reported Percentage	10.29%	7.35%	2.94%	7.35%	26.47%	1.47%	8.82%	35.29%
5	Estimated Percentage	15.87%	13.65%	12.38%	16.51%	5.71%	1.90%	2.22%	31.75%
	Reported Percentage	24.24%	24.24%	19.70%	6.06%	0.00%	0.00%	0.00%	25.75%
6	Estimated Percentage	15.55%	13.02%	9.84%	20.32%	5.40%	2.22%	2.22%	31.43%
	Reported Percentage	11.94%	16.42%	28.36%	20.90%	2.99%	1.49%	1.49%	16.42%
7	Estimated Percentage	15.24%	12.38%	8.25%	14.29%	13.33%	2.86%	2.86%	30.79%
	Reported Percentage	10.61%	18.18%	13.64%	9.09%	19.70%	3.03%	4.55%	21.21%
8	Estimated Percentage	13.97%	11.43%	6.98%	12.70%	20.63%	2.86%	3.17%	28.25%
	Reported Percentage	0.00%	7.46%	8.96%	2.99%	40.30%	1.49%	5.97%	32.84%
9	Estimated Percentage	10.16%	10.48%	9.84%	32.38%	9.84%	6.03%	0.32%	20.95%
	Reported Percentage	1.54%	13.85%	26.15%	47.69%	4.62%	3.08%	1.54%	1.54%
10	Estimated Percentage	8.25%	9.52%	9.21%	32.70%	15.87%	6.03%	0.32%	18.10%
	Reported Percentage	6.25%	12.50%	15.62%	31.25%	23.44%	3.12%	3.12%	4.69%
11	Estimated Percentage	6.35%	8.25%	7.30%	20.95%	29.52%	6.67%	2.54%	18.41%
	Reported Percentage	5.97%	7.46%	8.96%	7.46%	41.79%	5.97%	7.46%	14.93%
12	Estimated Percentage	4.76%	6.67%	6.03%	23.17%	33.65%	7.62%	2.86%	15.24%
	Reported Percentage	1.49%	4.48%	4.48%	4.48%	61.19%	5.97%	5.97%	11.94%
13	Estimated Percentage	10.16%	10.48%	10.16%	31.43%	10.48%	6.67%	0.32%	20.32%
	Reported Percentage	0.00%	4.62%	15.38%	72.31%	1.54%	1.54%	4.62%	0.00%
14	Estimated Percentage	8.25%	9.52%	9.52%	31.75%	16.51%	6.67%	0.32%	17.46%
	Reported Percentage	0.00%	15.15%	10.61%	45.45%	19.70%	1.52%	3.03%	4.55%
15	Estimated Percentage	6.35%	8.25%	7.62%	20.00%	30.16%	7.30%	2.54%	17.78%
	Reported Percentage	0.00%	7.46%	2.99%	8.96%	65.67%	2.99%	2.99%	8.96%
16	Estimated Percentage	4.76%	6.67%	6.35%	22.22%	34.29%	8.25%	2.86%	14.60%
	Reported Percentage	1.52%	3.03%	1.52%	15.15%	54.55%	6.06%	9.09%	9.09%
17	Estimated Percentage	12.38%	12.06%	14.92%	23.81%	6.98%	5.40%	0.32%	24.13%
	Reported Percentage	2.99%	25.37%	35.82%	19.40%	5.97%	0.00%	2.99%	7.46%
18	Estimated Percentage	11.11%	11.11%	13.65%	22.22%	14.29%	5.40%	0.63%	21.59%
	Reported Percentage	0.00%	10.61%	10.61%	16.67%	43.94%	1.52%	6.06%	10.61%
19	Estimated Percentage	12.06%	11.43%	12.38%	27.62%	6.67%	5.71%	0.32%	23.81%
	Reported Percentage	5.88%	13.24%	32.35%	33.82%	8.82%	0.00%	1.47%	4.41%
20	Estimated Percentage	10.79%	10.48%	11.11%	26.03%	13.97%	5.71%	0.63%	21.27%
	Reported Percentage	0.00%	11.94%	7.46%	22.39%	44.78%	1.49%	5.97%	5.97%
21	Estimated Percentage	13.97%	12.38%	11.75%	16.83%	8.57%	2.54%	2.54%	31.43%
	Reported Percentage	4.55%	9.09%	12.12%	4.55%	21.21%	3.03%	12.12%	33.33%
22	Estimated Percentage	12.38%	10.79%	10.48%	19.05%	12.70%	3.49%	2.86%	28.25%
	Reported Percentage	4.55%	7.58%	3.03%	3.03%	28.79%	1.52%	4.55%	46.97%

3.5.4. Results

Table 9 presents the estimated (Pre-Experiment) and subjectively reported (Preliminary Experiment) Valence, Arousal and Dominance levels for each sub-game. The estimated values are calculated by adding the incident's (VR parameter) affective values (according to "Interactive and Additive Effect" presented in Section 3.4.1); therefore the scaling is different from the sub-games' measured ratings, which are scaled between -3 and +3. Table 10 presents the estimated and reported occurrence probability (*OP*) of the each categorical label in each sub-game.

3.6. Conclusion

In this chapter an Affective VR, capable of evoking multiple emotional experiences on the users, was conceptualised and designed. The designed dynamic and interactive virtual environment can be reconfigured, by setting eight in-game parameters, to create up to 792 different sub-games (Sections 3.1, 3.2 and 3.3). By conducting an online survey the potential emotional effects of these parameters were estimated (Section 3.4). Then by employing an estimation algorithm, the overall potential emotional effects of all sub-games (combination of a number of in-game parameters) were also predicted (Section 3.4.1). By conducting a subjective experiment, the emotional experience of 68 participants, when exposed the 22 most powerful combinations (most affective sub-games – identified by the Cosine Similarity algorithm), were subjectively evaluated (Section 3.5). By designing the Affective VR, the emotional stimulation medium required for construction of the VR-based psychophysiological database is provided (research question II presented in Section 1.8).

The most important contribution of this chapter is the novelty of the designed Affective VR. As discussed in Section 1.7, majority of the designed 3D VRs (Wu et al., 2010; Parnandi et al., 2013; Rodríguez et al., 2015) or modified Retro Games (Liu et al., 2009; Chanel et al., 2011; Reuderink et al., 2013) attempted to influence either a single dimensional axis (either Valence or Arousal level) (Wu et al., 2010; Reuderink et al., 2013; Parnandi et al., 2013) or a single emotion label (e.g. evoking anxiety or depression) (Liu et al., 2009; Rodríguez et al., 2015). Only Chanel et al managed to influence their participants' Valence and Arousal levels (Chanel et al., 2011). However, the simplicity of their 2-dimensional Retro Game (Tetris) did not exploit the full potential of the virtual environments. The design of the Affective VR in this study, on the other hand, is a fully graphical (3D), dynamic and interactive (using both normal and force-feedback based controllers) virtual environment, capable of evoking multiple emotions on the users. Moreover, the ability to tune the settings of the VR easily to generate up to 792 different sub-games, with estimable emotional effects, makes this Affective VR extremely unique in the affective recognition field. From the literature, it should be remembered that various evaluated affective datasets have been presented in the form of images, sounds, music and video clips, to evoke various emotional experiences. To date, no form of Affective VR has been presented.

Chapter 4
Analyses of Emotional Experiences Within The Affective VR

Abstract – The evaluation of the emotional stimulation effectiveness of the Affective VR is described in this chapter, prior to any implementation within a physiological measurement paradigm. The analysis of the estimated (Pre-Experiment) and subjectively evaluated (Preliminary Experiment) affective powers of the VR variations (Sub-Games) provides an early level of confidence that this particular Affective VR is capable of manipulating individuals' emotional experiences, through the control of its internal parameters. Moreover, the approximation technique proved to be fairly reliable in predicting users' potential emotional responses, in various Affective VR settings, prior to actual experiences. Finally, the analysis suggests that the emotional responses of the users, with different age, gender and gaming experiences, could vary when presented with the same Affective VR situation. It was highlighted that the participants, aged between 18 and 30, experience similar emotional patterns, within a single affective session, in contrast to the patterns demonstrated by others in the 30 to 40 age group. Furthermore it was concluded that the female non-gamers' emotional experience highly deviates from other groups, when exposed to a single affective session.

4. Analyses of Emotional Experiences Within the Affective VR

4.1. Emotional Experience Forecasting

Theoretically, there is a debate within the literature surrounding the ability to accurately forecast an emotional response to future events. The prediction of future emotional reactions of oneself and others influence decision and behaviours in different circumstances (Van Kleef, 2009). Therefore emotional prediction plays an important role in human social life. Paradigms that assess a participant’s ability to predict their emotional responses typically are described verbally (such as the responses to the survey conducted in Section 3.4), while responses to current events are estimated based on the actual experience of the event.

Affective forecasting is performed in 2 different ways: based on imagination or prior experience. The two modes are not mutually exclusive. Imagination-based forecasting means that a person imagines himself or herself in a situation and assesses the potential emotional outcome. Experience-based forecasting involves the recall of similar past events, to evaluate future reactions. Robinson and Clore suggested that simulation and recalling of affective experiences are different in nature; however, they are similar in a sense, in that, in both conditions, the person is not actually present at the emotional situation (Robinson & Clore, 2001). Nevertheless, regardless of the selected technique (simulation or recall), assessment or appraisal of the affective situation plays an important role in the process. Fernandez-Duque and Landers argue that people are relatively skilled in predicting the type (e.g. positive/negative or the category) of the emotional experiences prior to a particular event; however, they are less skilled at estimating the intensity (e.g. levels of arousal or valence, or of an emotion label) and duration of the affective experiences (Fernandez-Duque & Landers, 2008).

Different studies have attempted to reveal and test the accuracy of human affective prediction, for future emotional events. As an illustration, Robinson and Clore verbally described 10 images from International Affective Picture System (IAPS) to participants and asked them to report their emotional experiences. This was undertaken by rating the intensity of 20 Emotion Labels, within 7 distinct clusters. Then the 10 images were presented to the participants to rate their emotional experience using the same procedure. They found that imagined emotions, based on verbal descriptions, are nearly identical to the affective reactions elicited when viewing these images (Robinson & Clore, 2001). Fernandez-Duque and Landers tested the ability to forecast an emotional response using a loss-reward game. The game was verbally described to the participants who were asked to estimate their emotional response (specifically regret at the choices they made) to a possible winning or losing outcome. They next played the game and were asked to report their emotional reactions following the actual experience of the game. The authors found that the experienced and predicted emotions, reported by those who suffered a wide-margin loss, were highly similar, whereas those with narrow-margin losses, experienced emotions with higher intensities than those they had imagined earlier (Fernandez-Duque & Landers, 2008).

Combining the data of the Pre-Experiments (Section 3.4.1) and the Preliminary Experiment (Section 3.4.2) enables one to assess the ability of participants in predicting their emotional responses to game scenarios. In Pre-Experiments, participants estimated their emotional responses to verbal descriptions of an in-game incident. Using an additive model, this provides an overall estimated emotional response to the sub-games, created by combining several incidents. In Preliminary Experiment, different sets of participants played 22 affective sub-games and provided a report of their emotional experience (Section 3.5). The analysis of the current data aimed to investigate the following two subjects:

1. **Affective Forecasting Accuracy of the Entire Sub-Games:** We evaluated the precision of the affective forecasting algorithm (Section 3.4.1), in predicting the affective power of the entire sub-games.
2. **Affective Models Comparison:** We employed different affective models (dimensional *vs*. categorical), to compare their capabilities and accuracy.

4.1.1. Results

To compare the estimated *vs*. reported emotions of the sub-games (using both dimensional and categorical models of affect), Table 9 and Table 10 (presented in Section 3.5.4) have been analysed.

Dimensional Model: Figure 14 represents the estimated and reported emotions along the three dimensions. The scores of each dimension were normalised (using Z-Score normalisation technique – 5.1.4) to enable a numerical comparison. As can be seen, the estimated emotions (blue triangles), along all three axes showed smaller variations relative to the reported emotions (red circles), although both estimated and reported had a similar directional change. Moreover, similar directional changes are also evident in the reliable strong correlations between estimated and reported, for all the three emotional dimensions (Table 11). This is in line with previous observation showing a good ability to predict the type but a poor ability to predict the intensity of the emotional experiences. To better understand the relationship between predicted and reported emotions, we next computed the difference between the two scores (estimated value subtracted from the reported level – Figure 15 and Table 12). As can be seen, the participants consistently underestimated the intensity of the Valence and Arousal levels; while the estimation of Dominance showed a large degree of variability across the sub-games and did not show a consistent under- or over-estimation (Table 12). To assess the accuracy of the emotion estimation technique even further, the estimated and reported emotional experiences of the participants, within each sub-game, have been presented by 2-dimensional vectors representing Valence and Arousal levels (Circumplex of Affect – Section 2.3.5). Then the direction of the estimated and reported vectors (*colour of the emotions*) has been compared, using the Cosine Similarity Algorithm (Pang-Ning Tan, 2005) (refer to Appendix D). Moreover, the magnitude of the estimated and reported vectors (*intensity of the emotions*) has also been subtracted, to highlight the absolute intensity differences between the estimated and reported emotions. Figure 17 represents the cosine similarity and absolute magnitude difference ranges, across all 22 sub-games. As can be seen in Figure 16, 70% of the estimated emotional experiences are more than 80% similar (cosine similarity level) to the corresponding reported emotional experiences. However, the mean absolute magnitude difference (emotional intensity difference) across sub-games has been -0.93 (±0.46 – significantly different from zero, P<0.001), representing a significant under-estimation.

Categorical Model: The correlation between estimated and reported Occurrence Probability (OP) of the Emotion Labels varied between the labels (Table 11). There was a strong correlation for Excited, Angry and Content; a moderate correlation for Relaxed, Happy, Afraid and Bored; and a weak and unreliable correlation for Sad. As all correlations were positive, it again demonstrates that there was a good prediction of the type of emotion that would be experienced. The difference between estimated and reported labels are reported in Figure 16, with the reliability test in Table 12. Participants in the survey reliably used Relaxed, Afraid and Bored labels more frequently when classifying their emotions, in comparison to those who played the sub-games. In contrast, the participants selected Angry and Sad labels less frequently in the survey, compared to when playing the games. There was a trend for an under-estimation of the Happy label, with no reliable difference between Content and Excited.

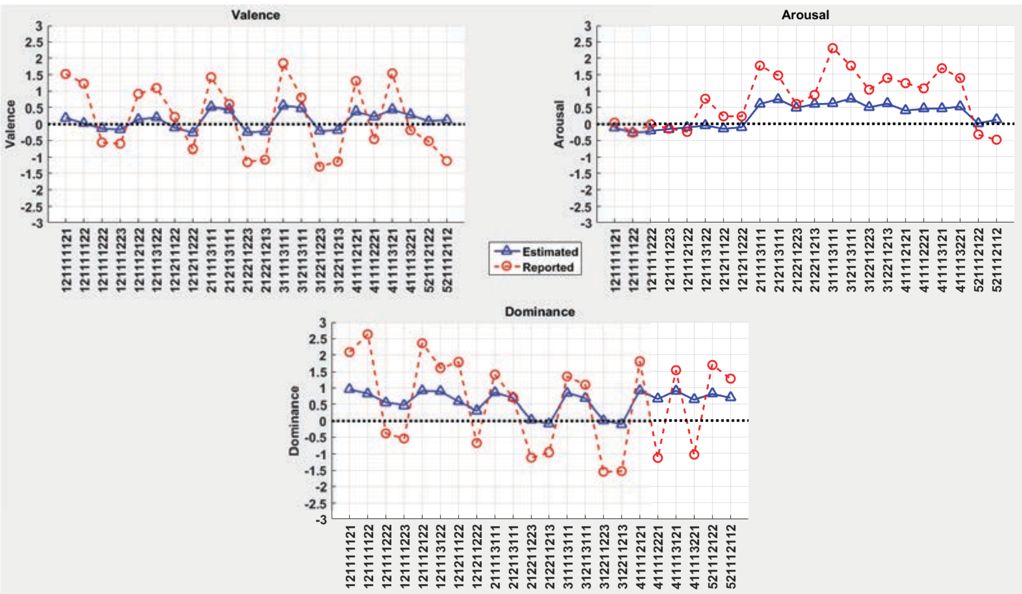


Figure 14 – Estimated *vs*. Reported Emotional Powers of 22 Sub-Games – The Blue-Solid-Triangles and Red-Dashed-Circles Represents (Respectively) The Estimated and Reported Affective Powers, Within Each Axis

Table 11 – Estimated *vs*. Reported Emotional Power and Occurrence Percentage Correlation Coefficients and Significance Level, Within Dimensional and Categorical Spaces (Respectively)

Axis	Correlation Coefficient – r(22)	P-Value
Valence	0.775	<0.001
Arousal	0.855	<0.001
Dominance	0.829	<0.001
Mean Across Axes	**0.820 (±0.04)**	**<0.001**
Relaxed	0.689	<0.001
Content	0.713	<0.001
Happy	0.576	0.005
Excited	0.839	<0.001
Angry	0.885	<0.001
Afraid	0.562	0.006
Sad	0.390	0.072
Bored	0.595	0.003
Mean Across Emotion Labels	**0.656 (±0.16)**	**0.011 (±0.024)**

Table 12 – Estimated *vs*. Reported Emotional Power and Occurrence Percentage Difference Mean and Range, Within Dimensional and Categorical Spaces (Respectively) – Estimated Values are Subtracted From the Reported Values – results of one sample T-Test Comparing the Differences' Mean and Zero

Axis Emotion Label	Difference Mean (25^{th} – 75^{th} Percentiles)	Stat for Mean Significantly Different From Zero – t(df)	P-Value (T-Test) for Mean Significantly Different From Zero
Valence	-0.51 (-0.95 – -0.31)	t(21) = -3.4535	0.002
Arousal	-0.47 (-0.83 – -0.11)	t(21) = -3.9903	<0.001
Dominance	-0.18 (-0.88 – 0.98)	t(21) = -0.7527	0.459
Relaxed	4.88% (3.24% – 9.39%)	t(21) = 3.5807	0.001
Content	-2.25% (-5.62% – 3.21%)	t(21) = -1.5740	0.130
Happy	-4.63% (-7.31% – 3.04%)	t(21) = -2.4046	0.025
Excited	3.35% (1.44% – 11.04%)	t(21) = 1.1999	0.243
Angry	-9.97% (-19.66% – 2.41%)	t(21) = -3.5335	0.001
Afraid	2.52% (1.06% – 4.22%)	t(21) = 6.4028	<0.001
Sad	-2.71% (-4.92% – -1.15%)	t(21) = -4.3588	<0.001
Bored	8.82% (3.29% – 16.66%)	t(21) = 3.9731	<0.001

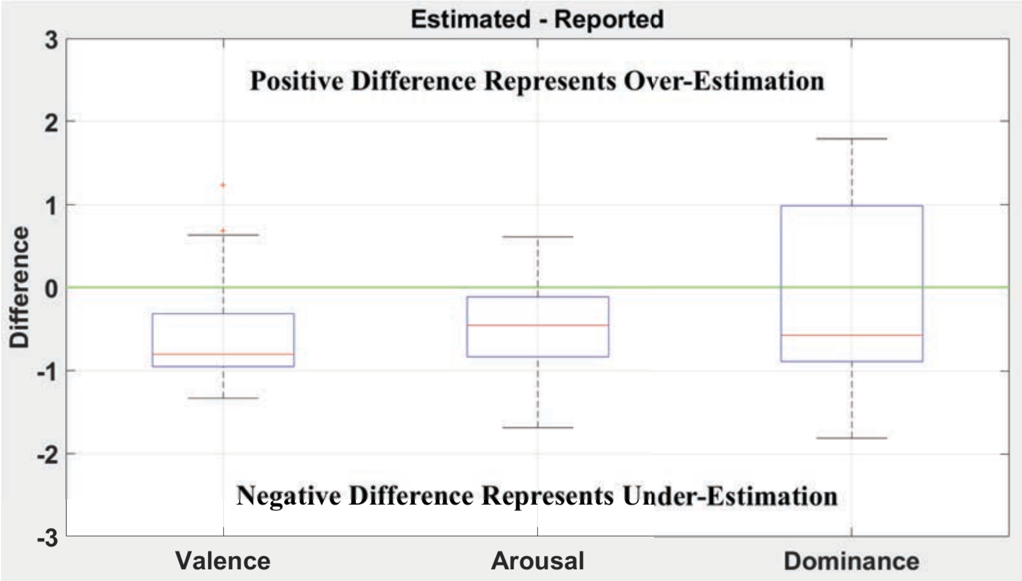


Figure 15 – Estimated *vs*. Reported Dimensional Power Difference – Positive and Negative Differences Represent, Respectively, Over and Under Power Estimation – The Green Horizontal Line Represents the Zero-Difference Level – The Blue Boxes Represent the 25th and 75th Data Points Intervals – The Red Lines Represent the Median Points

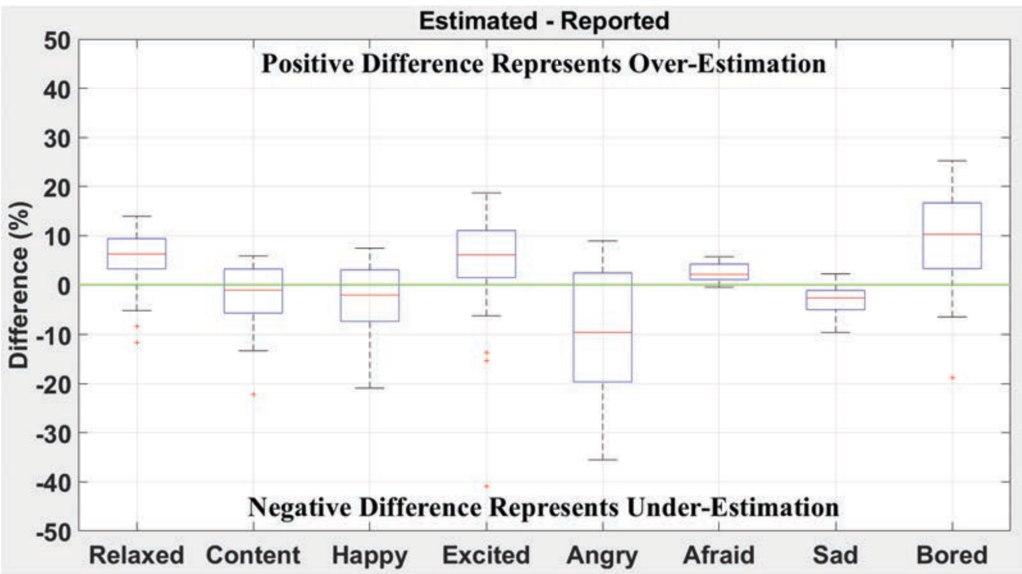


Figure 16 – Estimated *vs*. Reported Categorical Occurrence Percentage Difference – Positive and Negative Differences Represent, Respectively, Over and Under-Estimation – The Green Horizontal Line Represents the Zero-Difference Level – The Blue Boxes Represent the 25th and 75th Data Points Intervals – The Red Lines Represent the Median Points

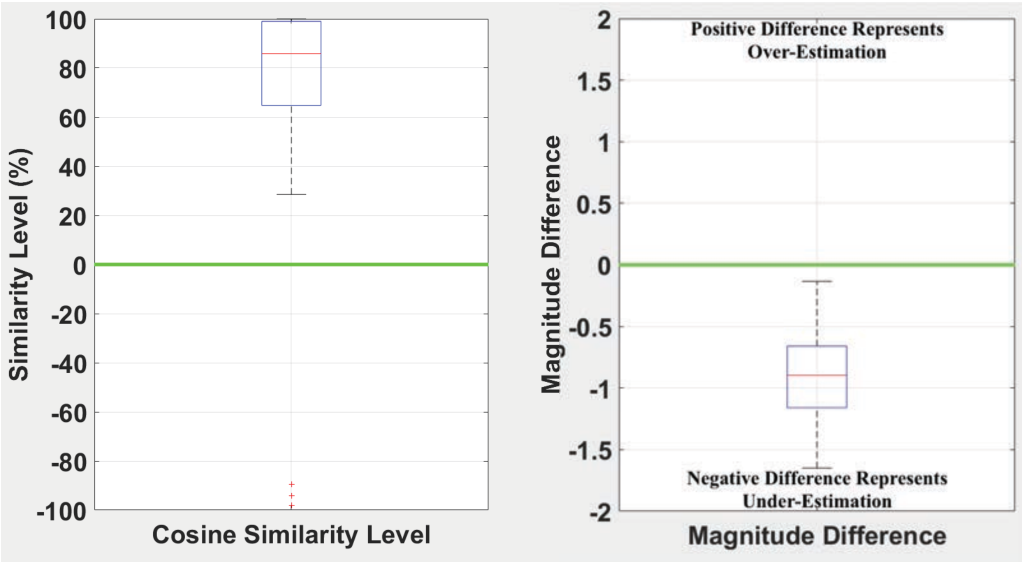


Figure 17 – Estimated vs. Reported Emotion Comparison, Within the Circumplex of Affect – The Blue Boxes Represent the 25th and 75th Data Points Intervals – The Red Lines Represent the Median Points

4.1.2. Discussion

As shown above, there was a high positive and significant correlation between the estimated and reported Valence, Arousal and Dominance levels of the 22 sub-games. This shows that (as an illustration), if a particular sub-game has been estimated to have a high positive arousing effect, the participants, on average, reported a high positive arousing experience, and vice versa (similar pattern in all axes). Moreover, the majority of the estimated emotional experiences in the Circumplex of Affect were highly similar, when compared to their corresponding, subjectively reported emotional experiences (high cosine similarity level – similar colour of emotion). However, there was a significant underestimation in Valence, Arousal and intensity of the emotions. Therefore, the results suggest that, although the estimation algorithm introduced in Section 3.4.3, was (fairly) accurate in estimating the colour of the participants' emotional experiences (high correlation and similarity levels), their intensity has always been under-estimated (on average, a 0.5 under-estimation in Valence and Arousal levels, and a 0.93 under-estimation in the emotional experience intensity). This pattern is similar to what has been suggested in the literature (Fernandez-Duque & Landers, 2008). Finally, as the mean of the Dominance difference is not significantly different from zero, it has been concluded that the Dominance levels have been almost equally under- and over-estimated, and this finding did not support any clear conclusion.

Similar to the dimensional analysis, high and significant correlation patterns between the estimated and reported occurrence percentages of the Emotion Labels have been observed. But the results suggest that, although the emotion estimation algorithm was able to (fairly) accurately estimate the tendency of the occurrence percentages of the Emotion Labels, Relaxed, Afraid and Bored labels were significantly over-estimated, while Happy, Angry and Sad labels were significantly under-estimated. Estimated *vs.* reported occurrence percentage differences of the

Content and Excited labels were not significantly different from zero. Therefore, it has been concluded that these two labels have been almost equally under- and overestimated, and, as before, was unable to support a clear conclusion.

Finally, as can be observed in the results, the correlation between estimated and reported Valence, Arousal and Dominance levels, is not only higher, but also slightly more significant, when compared to the same analysis within the categorical model (Table 11). One can conclude that the dimensional affective space has been more accurate (when compared to the categorical model) in estimating the emotional experience of the participants, within any sub-game. Moreover, the dimensional space demonstrated a consistent under-estimation pattern, in approximating the emotional experience of the participants (in Valence and Arousal levels, and emotional experience intensity). In contrast, the categorical model developed both under and over-estimated predictions, in different Emotion Labels. Therefore, we can conclude that although both dimensional and categorical models could represent the affect space, the dimensional model provides a more reliable platform for estimating the emotional experience of the participants (accurate colour while reliable under-estimating the intensity of the emotion), when compared to the categorical model.

4.2. Affective VR Emotional Stimulation Efficiency

As discussed in Chapter 3, the designed Affective VR evoked multiple emotional experiences in the participants as a result of changes in its internal parameters (incidents – Section 3.2). Moreover, it was suggested that the designed Affective VR evoked emotions across the entire Circumplex of Affect (Section 2.3.5). However, in the context of an active game, it was a challenge to evoke emotions that match in intensity to those typically reported in the low arousal high pleasure quadrant. Figure 18 presents the distribution of the reported emotions of the 22 sub-games, within the Circumplex of Affect. Each dot represents the mean Valence *vs.* Arousal subjective ratings (across participants) of a particular sub-game. As can be seen in the graph, on average, the Affective VR variations (22 sub-games) evoked a wide range of emotions, occupying the entire 2-dimensional Circumplex of Affect, although dominating the upper space. This is represented as a robust main effect of sub-game in the following analysis, where participants are treated as random variables. An Analysis of Variance (ANOVA)[19] revealed a significant difference among the ratings of sub-games, with respect to each affective axis (Valence, Arousal and Dominance – $P<0.001$). Moreover, 2D[20] and 3D[21] Multi-variant Analysis of variance (MANOVA) highlighted a significant difference between the dimensional ratings of the sub-games ($P<0.001$). On the other hand, Multi-variant Analysis of variance (MANOVA) based on the Emotion Labels[22] revealed that there is a significant difference among the selected Emotion Labels across the sub-games

[19] One of the axis (Valence, Arousal and Dominance) considered as the dependent variable, while the game code as the independent parameter.
[20] Valence and Arousal considers as dependent variables, while the game code as the independent parameter.
[21] Valence, Arousal and Dominance considers as dependent variables, while the game code as the independent parameters.
[22] The dimensions of an 8-dimensional binary vector, representing the selected Emotion Label within each sub-game, were considered as the dependent variables, while the game code as the independent parameter.

(P<0.001). From these findings, one can conclude that the single controllable Affective VR, designed in this study, has been able to effectively manipulate the participants' emotions, by controlling its internal parameters (incidents).

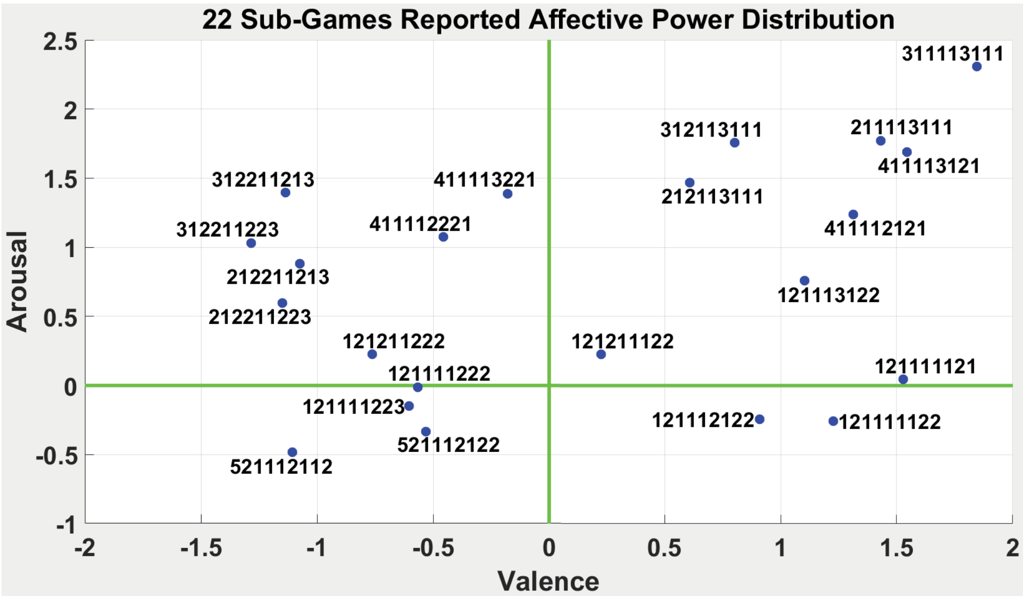


Figure 18 – Subjective Distribution of 22 Sub-Games Within the 2-Dimensional (Valence *vs*. Arousal) Circumplex of Affect – Each Dot Represents the Mean Valence *vs*. Arousal Subjective Ratings (Across Participants) of a Particular Sub-Game

4.3. Effects of Individual Differences in Emotional Experience

Despite the effort to minimise the variability between participants, in each individual affective session (sub-game), whilst maximising the variability between sessions' experiences (Section 3.1), emotional experience inconsistencies amongst participants, within a single affective sub-game, were observed. In 2003, Eugène et al. suggesting that, in the area of emotion recognition and affective computing, the individual differences should be considered as the rule rather than the exception (Eugène et al., 2003). In a given affective situation, different individuals can respond differently, thereby causing too much variability within the affective database (Hamann, 2004). Therefore, in the next section, we examined four potential sources of these individual differences.

4.3.1. Sources of Individual Differences

According to Hamann, individual differences can be categorised into (1) **Personality**, (2) **Gender** and (3) **Experience** or Affective Skills, each of which differentiating the participants' emotional experiences, when exposed to a single emotional situation (Hamann, 2004). (4) **Age**, on the other hand, has also been identified in various literature sources, as another factor in the differentiation of individual affective experiences (Gross et al., 1997; Mroczek, 2001; Combain et al., 2004; Schubert, 2007).

4.3.1.1. Personality

In 1985, Eysenck introduced two emotional traits described as **Extraversion** and **Neuroticism**. They believed that individuals classed as extroverts have a higher susceptibility in experiencing positive affect, while neurotic characters have a higher susceptibility in experiencing negative affects (Eysenck & Eysenck, 1985). As an illustration, in an uncertain affective situation (e.g. when one notices that one's boss is watching), an extrovert is more likely to experience a positive emotion (e.g. "someone has notice my hard work"), whereas, a neurotic individual is more likely to feel a negative emotion (e.g. feeling nervous and insecure) (Zelenski & Larsen, 1999). In 1991, Larsen and Ketelaar investigated the relationship between the theory, introduced by Eysenck, and affective experiences. They presented a number of affective texts to a group of participants and asked them to rate their emotional experiences, using 12 affective labels (e.g. Sad, Happy, etc.). They concluded that different personalities are prepared to respond with specific and different emotions, when presented with the same stimuli. They further suggested that extroverted individuals are prepared to respond with stronger positive emotions rather than negative emotions; while neurotic persons are more likely to experience negative emotions rather than positive emotions, given similar affective circumstances (Larsen & Ketelaar, 1991).

4.3.1.2. Gender

To date, many studies have investigated the emotional differences between males and females. In 1993, Grossman and Wood performed two separate experiments. In the first experiment, they asked participants to report the frequency and intensity of their day-to-day emotions. They concluded that women report more intense and more frequent emotional experiences than men. In the second experiment, they presented a number of images from the International Affective Picture System (IAPS) collection to a number of participants and asked them to report how positive or negative (pleasured/displeasured) they felt after each image. They also measured the participants' heart rates and EMG signals, while viewing the affective images. They concluded that females report more intense emotions and also generate more extreme physiological reactions than males (Grossman & Wood, 1993). Moreover, Karing and Gordon investigated this difference, by showing affective video clips to a number of participants, while recording their GSR and facial expressions (through image processing). They also employed the Valence/Arousal affective space, plus a 4-label emotional questionnaire as their subjective affective assessment. They concluded that women are more expressive than men. However, they do not experience significantly stronger emotion, compared to men (Kring & Gordon, 1998). Furthermore, Bradley et al. used IAPS images to evoke a number of emotional states on a number of participants, while recording their GSR, heart rate and EMG responses. The 3-dimensional Valence/Arousal/Dominance affective space, together with a 24-lable emotional questionnaire, was also employed as their subjective affective measure. They found a level of agreement among the male and female subjective ratings. However, in different picture clusters, they identified a significant difference between male and female in both subjective ratings and physiological responses (Bradley et al., 2001). In 2001, Wild et al. performed a similar experiment and presented a number of affective images to a group of participants. They asked them to rate their emotional experiences, using Valence and a 6-lable emotional questionnaire. They concluded that females report positive images as more pleasant and negative pictures as more unpleasant, when compared to males (Wild et al., 2001).

4.3.1.3. Experience or Affective Skills

The emotional response, in an affective situation, can vary according to the individual's emotional regulation strategy and expertise. As an illustration, the emotional experience of an experienced rugby or cricket audience would be different, compared to an inexperienced audience, comprising those who neither know the rules, nor have had any past exposure to the games; although, both may experience the same circumstances, in a single game. In 2005, Bigand et al. examined this relationship, by playing a number of music tracks to participants with and without musical training. They defined the affective skills based on academic musical training, and selected them from graduate music students. They instructed the participants to subjectively assess the emotional effect of each excerpt, according to a number of emotional labels, at the end of the experiment. Participants were allowed to listen to the musical tracks as many times as they wished. They concluded that affective responses to music do not strongly depend on the level of musical training or experience (Bigand et al., 2005). On the other hand, Kantor-martynuska & Horabik repeated the same experiment (using the same music employed by (Bigand et al., 2005)), while employing different definitions for "experience" (someone who is capable of playing an instrument). In contrast to (Bigand et al., 2005), they identified a strong positive correlation between emotional responses and musical expertise level (Kantor-martynuska & Horabik, 2015). In 2009 Pollak et al. performed an experiment to compare the emotional response of children to a number of affective facial expressions, presenting five basic emotions (Scared, Sad, Happy, Surprised, and Angry). They employed normal children and a group of abused children, who had been exposed to high levels of parental anger. They conclude that the two groups did not differ in recognising Scared, Sad, Happy and Surprised emotions, whereas abused children were able to recognise Angry expressions earlier than normal children (Pollak et al., 2009). In 2011, Silvia & Berg examined the "expertise effect" on emotional responses in movies, by presenting a number of film clips to a number of participants with different level of expertise (critics *vs*. normal audience). They concluded that experts found movies more interesting and less confusing, compared to inexperienced individuals. They claimed that the experts' level of interest could be predicted by the movie complexity, more accurately than inexperienced participants (Silvia & Berg, 2011).

4.3.1.4. Age

In 1997, Gross et al. conducted a survey about the intensity and control level of the daily emotional experiences of 258 participants with various age, ethnical and cultural diversities. They concluded that age is a significant factor in creating different levels of intensity and control of daily emotional experiences. They stated that older participants reported fewer negative emotional experiences and higher emotional control, when compared to younger people (Gross et al., 1997). In 2001, Mroczek analysed the *Midlife in the United States* (MIDUS) survey, and concluded that older people (compared to younger individuals) report stronger positive affects, while experienced weaker negative emotions (Mroczek, 2001). Moreover, Schubert employed 5 western romantic pieces of music, and compared the emotional responses of 65 participants, with various ages. They concluded that older people tend to exhibit more positive emotions, with generally lower emotional intensity, compared to younger individuals (Schubert, 2007).

4.3.2. Results

In the present study, the participants' gender, age and gaming experience (see the gamer *vs*. non-gamer definition in Section 3.4.2) were recorded before each experiment (refer to Appendix E). However, as the personality of the participants was not assessed as part of the experimental design, we were unable to analyse the personality effect on the emotional experiences. As discussed earlier (Section 3.5.3), the participants consisted of 4 different groups; male gamers, female gamers, male non-gamers and female non-gamers (17 participants for each group). Moreover, the age of the participants was classified into 4 classes (12-18, 18-24, 24-30 and 30-40 years old). Unlike the gender and gaming experience groups, the participants within the age groups are not evenly distributed. Overall 40 participants belonged to the 18-24 age group, 20 to the 24-30 age group and only 8 to the 30-40 age group. No participant, aged between 12 and 18, was recruited to take part in this experiment.

A 2D[23] and 3D[24] Multi-variant Analysis of Variance (Table 13) highlighted a significant difference between the dimensional ratings of the participants, with various gaming experiences (gamers *vs*. non-gamers) and ages (18-24 *vs*. 24-30 *vs*. 30-40). However the same analysis revealed a marginally significant gender difference between male and female participants in the 2D Circumplex of Affect, while there were **no** significant differences among participants with various genders in 3D Affective Space. Moreover, the analysis revealed a significant interaction effect between experience, gender and age (except gender *vs*. age in both 2D and 3D and gender *vs*. experience in 2D affective space – refer to Table 13). Therefore we concluded that experience, age and gender could be considered as significant variables in creating individual differences in emotional experiences (in both 2D and 3D affective spaces), given a particular affective stimulus[25].

---

[23] Valence and Arousal considers as dependent variables, while the game code, experience, gender and age as the independent parameters.

[24] Valence, Arousal and Dominance considers as dependent variables, while the game code, experience, gender and age as the independent parameters.

[25] In both 2D and 3D MANOVA analyses, the game code has been identified as a significant factor, as discussed in Section 4.2.

Table 13 – Game Code, Experience, Gender and Age Effects on 2D and 3D Dimensional Affective Space (MANOVA Analysis) – [a], [b] and [c] Represent Significant, Marginally Significant and Insignificant Result, Respectively

Single/Interactive Factor	**2D MANOVA P-Value**	**3D MANOVA P-Value**
Game Code	<0.001 [a]	<0.001 [a]
Experience	0.026 [a]	0.026 [a]
Gender	0.055 [b]	0.121 [c]
Age	<0.001 [a]	<0.001 [a]
Experience * Gender	0.129 [c]	0.063 [b]
Experience * Age	0.011 [a]	0.038 [a]
Gender * Age	0.426 [c]	0.290 [c]
Experience * Gender * Age	<0.001 [a]	<0.001 [a]

To be able to compare the emotional experiences of various groups with different age, gender and gaming experiences, two distinct analyses have been conducted, based on the calculation of (i) the absolute differences of the affective powers (dimensional model), occurrence percentages (categorical model), (ii) the cosine similarity (emotional colour comparison) and magnitude absolute differences (emotional intensity comparison) of emotional experiences. As the distribution of participants with different ages is not evenly allocated among participants with different genders and gaming experiences, the above analyses (i and ii) have been conducted separately on (a) different genders and gaming experiences (four classes: male gamer, male non-gamer, female gamer and female non-gamer) and (b) distinct ages (three classes: 18-24, 24-30 and 30-40). Table 14 presents the means and ranges of the absolute differences of the emotional powers and occurrence percentages, for various age comparisons. Moreover the table presents the means and ranges of cosine similarities and absolute magnitude differences of various age comparisons. As an illustration, the reported Valence levels of participants, aged 18-24 and 24-30, are (in average) 0.38 different (either higher or lower) across all 22 games (the absolute difference is calculated to reveal the total intensity difference). Whereas, the differences between the reported Valence levels of participants, aged 30 to 40 years, are much larger when compared to 18-24 or 24-30 year-old individuals (0.88 and 0.99 respectively). Table 15 presents the same analyses, across participants with different genders and gaming experiences (male gamers, female gamers, male non-gamers and female non-gamers).

Table 14 – Emotional Power and Occurrence Percentage Absolute Difference Mean and Ranges, Within Dimensional and Categorical Spaces (Respectively), According to Different **Age** Comparisons – Cosine Similarity and Magnitude Absolute Difference Mean and Ranges, Within 2-Dimensional Affective Space – (A - B) Presents the (A) 25^{th} and (B) 75^{th} Percentiles

	Axis / Label	18-24 *vs.* 24-30	18-24 *vs.* 30-40	24-30 *vs.* 30-40
	Valence	0.38 (0.13 – 0.49)	0.88 (0.52 – 1.22)	0.99 (0.5 – 1.53)
	Arousal	0.24 (0.10 – 0.37)	0.75 (0.49 – 0.97)	0.71 (0.48 – 0.96)
	Dominance	0.34 (0.13 – 0.46)	0.7 (0.22 – 0.97)	0.53 (0.10 – 0.75)
	Relaxed OP (%)	5.8 (0 – 10)	7.07 (0 – 10.25)	4.75 (0 – 9.52)
	Content OP (%)	6.94 (3.25 – 9.29)	11.91 (7.5 – 17.56)	9.56 (5 – 15)
	Happy OP (%)	6.7 (2.5% – 10.23)	15.6 (5.8 – 20)	14.4 (3.80 – 25)
	Excited OP (%)	8.74 (2.5% – 12.73)	19.03 (7.80 – 22.92)	18.8 (5 – 25)
	Angry OP (%)	6.9 (3.04% – 11.66)	11.69 (4.87 – 17.5)	12.61 (0 – 20)
	Afraid OP (%)	3.23 (2.31% – 4.87)	2.13 (0 – 2.5)	2.88 (0 – 5)
	Sad OP (%)	4.84 (0.23% – 7.5)	5.4 (2.43 – 7.5)	4.65 (0 – 9.52)
	Bored OP (%)	6.39 (2.55% – 10.47)	15.08 (7.31 – 20)	10.53 (4.76 – 15)
2D	**Cosine Similarity (%)**	94.09 (94.71 – 99.87)	50.77 (36.65 – 98.51)	47.01 (15.2 – 99.02)
	Magnitude Difference	0.36 (0.16 – 0.44)	0.6 (0.24 – 1.08)	0.5 (0.11 – 0.77)
3D	**Cosine Similarity (%)**	96.88 (95.87 – 99.4)	66.26 (46.46 – 98.33)	61.97 (53.85 – 98.26)
	Magnitude Difference	0.45 (0.23 – 0.56)	0.63 (0.23 – 0.91)	0.43 (0.14 – 0.7)

Table 15 – Emotional Power and Occurrence Percentage Absolute Difference Mean and Ranges, Within Dimensional and Categorical Spaces (Respectively), According to Different **Groups** Comparisons (Male Gamer, Male Non-Gamer, Female Gamer and Female Non-Gamer) – Cosine Similarity and Magnitude Absolute Difference Mean and Ranges, Within 2-Dimensional Affective Space – (A–B) Presents the (A) 25th and (B) 75th Percentiles

	Axis / Label	**Male Gamer *vs.* Male Non-Gamer**	**Male Gamer *vs.* Female Gamer**	**Male Gamer *vs.* Female Non-Gamer**	**Male Non-Gamer *vs.* Female Gamer**	**Male Non-Gamer *vs.* Female Non-Gamer**	**Female Gamer *vs.* Female Non-Gamer**
	Valence	0.47 (0.23 – 0.72)	0.46 (0.23 – 0.58)	0.47 (0.29 – 0.65)	0.51 (0.26 – 0.61)	0.46 (0.13 – 0.56)	0.33 (0.11 – 0.47)
	Arousal	0.46 (0.11 – 0.65)	0.4 (0.17 – 0.64)	0.5 (0.29 – 0.7)	0.56 (0.29 – 0.79)	0.55 (0.26 – 0.81)	0.42 (0.27 – 0.64)
	Dominance	0.3 (0.1 – 0.53)	0.33 (0.11 – 0.47)	0.66 (0.2 – 0.88)	0.36 (0.07 – 0.59)	0.63 (0.41 – 0.84)	0.56 (0.32 – 0.68)
	Relaxed OP (%)	3.96 (0 – 6.25)	5.61 (0 – 5.88)	5.73 (0 – 6.61)	6.8 (0 – 11.76)	6.75 (0 – 11.76)	7.4 (0 – 12.5)
	Content OP (%)	6.63 (1.10 – 6.98)	8.82 (5.88 – 11.76)	12.28 (5.88 – 13.6)	7.73 (4.77 – 11.39)	11.3 (0.36 – 17.27)	11.38 (4.77 – 15.12)
	Happy OP (%)	11.93 (5.88 – 17.27)	10.69 (0 – 17.64)	11.96 (5.88 – 16.17)	8.12 (5.51 – 11.76)	9.5 (0.36 – 15.17)	10.83 (5.88 – 17.64)
	Excited OP (%)	8.87 (5.51 – 12.5)	8.02 (5.88 – 5.88)	8.67 (0.36 – 12.86)	11.31 (0.36 – 19.48)	11.24 (5.88 – 16.91)	5.93 (2.74 – 6.25)
	Angry OP (%)	11.77 (0.36 – 19.85)	8.82 (0 – 17.64)	11.18 (2.94 – 16.91)	10.18 (2.2 – 14.7)	12.06 (2.35 – 21.32)	8.67 (0.36 – 7.45)
	Afraid OP (%)	3.02 (0 – 5.88)	4.81 (0 – 5.88)	2.53 (0 – 5.88)	3.49 (0 –5.88)	2.84 (0 – 6.25)	4.1 (0 – 5.88)
	Sad OP (%)	5.38 (0 – 6.25)	4.81 (0 – 5.88)	5.7 (0 – 7.14)	5.88 (0.36 – 6.98)	6.77 (0 – 12.5)	4.06 (0 – 5.88)
	Bored OP (%)	10.57 (5.88 – 14.33)	9.89 (5.88 – 11.76)	10.3 (4.04 – 11.76)	9.84 (0.73 – 17.64)	9.14 (1.83 – 12.5)	9.66 (0.73 – 11.76)
2D	**Cosine Similarity (%)**	87.08 (92.03 – 99.66)	77.07 (82.81 – 98.61)	73.75 (79.83 – 98.04)	74.82 (73.40 – 98.63)	78.38 (71.37 – 99.75)	74.73 (86.82 – 98.59)
	Magnitude Difference	0.48 (0.18 – 0.74)	0.36 (0.17 – 0.51)	0.48 (0.25 – 0.58)	0.44 (0.23 – 0.57)	0.49 (0.16 – 0.9)	0.36 (0.18 – 0.5)
3D	**Cosine Similarity (%)**	94.23 (86.67 – 98.53)	93 (89.04 – 98.62)	74.47 (84.18 – 98.23)	90.98 (86.46 – 98.08)	77.21 (92.46 – 99.56)	78.47 (88 – 98.23)
	Magnitude Difference	0.45 (0.29 – 0.65)	0.38 (0.18 – 0.46)	0.63 (0.27 – 0.89)	0.47 (0.18 – 0.72)	0.63 (0.29 – 0.91)	0.49 (0.32 – 0.64)

4.3.3. Discussion

As shown above, participants' gender, age and gaming experience are significant factors in creating various emotional responses, in a particular affective stimulation. The results presented in Table 14 suggest that the Valence, Arousal and Dominance difference levels of 18-24 and 24-30 year-old participants, across the 22 sub-games, are lower when compared to 30-40 year-old individuals. The same pattern can be observed in the occurrence percentage differences of Emotion Labels, across the sub-games. Moreover, the Cosine Similarity technique highlighted a significantly higher emotional colour similarity, and a minor magnitude (emotional intensity) difference between the 18-24 and 24-30 groups, when compared to the individuals aged between 30 and 40. These results suggest that the emotional responses of 18-24 and 24-30 year-old participants are similar, in contrast to those of the 30-40 year-old individuals. Therefore, to minimise the differences among individuals' emotional responses, it was decided to recruit only 18 to 30 year-old participants for any further physiological experiments (Chapter 5 and 6).

On the other hand, the comparisons between participants with different genders and gaming experiences (Table 15) highlighted consistent difference levels in the occurrence percentages of Valance, Arousal and Emotion Labels. This means that, as an illustration, the reported Valence levels of any group (male gamer, male non-gamer, female gamer and female non-gamer) have been (on average, across the sub-games) 0.45 lower or higher than the other groups. This feature resulted in almost identical cosine similarity and magnitude difference levels in 2-dimensional affective space (except higher cosine similarity levels in male gamer *vs*. male non-gamer groups – refer to Table 15). The Dominance difference levels of the non-gamer females, on the other hand, have been significantly higher (in average across the sub-games) when compared to other groups (refer to Table 15). Therefore, conducting the same similarity analysis, in 3-dimensional affective space, resulted in significantly lower cosine similarity and higher magnitude difference levels, when comparing non-gamer females, with other groups (refer to Table 15). This means that the non-gamer females have the least overall emotional experience similarity, when compared to the others. Therefore, to minimise the between-participant differences, when exposed to a particular affective stimulation, it was decided to avoid recruiting non-gamer females for any further physiological experiments (Chapter 5 and 6).

4.4. Conclusion

In this chapter, the designed Affective VR has been evaluated, through subjective assessments of 68 participants, with different age, gender and gaming experiences. It was highlighted that the designed Affective VR is capable of evoking the emotions, which were estimated using an approximation technique (Section 4.1). Moreover it was concluded that the designed Affective VR is capable of covering the entire Circumplex of Affect, by evoking multiple emotional experiences (Section 4.2 – research question II presented in Section 1.8). Furthermore, the analysis highlighted a significant individual difference, among the emotional experience of the participants, with different age, gender and gaming experiences, when exposed to the same sub-game (Section 4.3). Conducting this evaluation provided the required confidence level about the capabilities of the designed Affective VR in influencing its users' emotional experiences effectively. It also provides insight into the potential differences between participants' emotions, when exposed to similar affective stimulation. The findings of this chapter have been employed to select the most powerful sub-games, which have the highest probability of evoking the required emotional experiences in the psychophysiological database construction process. Moreover, according to the identified individual differences, it was decided to recruit only gamers, aged between 18 and 30, to participate in the psychophysiological database recording experiment, in order to minimise the emotional differences between participants.

The most important contribution of this chapter is the identification of the significant individual differences among participants with different age, gender and gaming experiences. It was highlighted that the female non-gamers could have a significantly different emotional experience, within a single game, when compared to other male (gamer or non-gamer) and female gamer users. Moreover, it was concluded that all participants aged above 30 years old (regardless of their gender and gaming experience) could experience a completely different emotion while exposed to a single game, when compared to other participants, aged between 18 and 30. These findings, coupled with the emotion estimation technique presented, could provide invaluable insights about the potential emotional response of the users to a new game (in the gaming industry), prior to its distribution. A capability of being able to identify the potential users (their gender, age and gaming experience) could provide extremely beneficial commercial and marketing information for the diffusion and deployment of the designed games.

Chapter 5

Affective Recognition System Design

Abstract – Detecting emotional responses in multimedia environments is an academically and technologically challenging research issue. Over the past 20 years, researchers in the domain of “Affective Computing” have investigated various psychophysiological parameters, in order to detect and quantify a wide range of human affective states. In this chapter, we present a detailed literature review of over 30 affective computing studies, undertaken since 1993. All aspects of these studies (stimuli type, pre-processing, windowing, features, classification technique, etc.) have been reported in detail. Moreover, a psychophysiological database, cataloguing the EEG, GSR and heart rate of 30 participants, exposed to an affective virtual environment, has been constructed (Primary Experiment – Experiment 4). 743 features were extracted from the physiological signals. Then, by employing a feature selection technique, the dimensionality of the feature space was reduced to a smaller subset, containing only 30 features. Four classification techniques (KNN, SVM, Discriminant Analysis (DA) and Classification Tree) were employed to classify the affective psychophysiological database into four Affective Clusters (derived from a Valence-Arousal space) and eight Emotion Labels. By employing cross-validation techniques, the performances of more than a quarter of a million different classification settings (under various pre-processing and classification settings) were investigated. The results suggested that the physiological signals could be employed to classify emotional experiences, with high precision. The KNN and SVM outperformed both Classification Tree and DA classifiers; with 97.01% and 92.84% mean accuracies, respectively.

5. Affective Recognition System Designing

In Chapter 3 and 4 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 chapter, by employing the designed Affective VR, an affective computing system is conceptualised, designed and evaluated. To do this, the relationship between psychophysiological signals and human emotions, evoked through the designed Affective VR, has been the focus of investigation. To support this research:

1. A detailed literature review over 30 affective recognition studies, published since 1993, has been conducted. (Section 5.1 of the present chapter).
2. A psychophysiological experiment has been conducted, in which simple EEG, GSR and Heart Rate signals of 30 participants have been recorded, whilst playing the most affective games. These physiological signals provided a comprehensive database for further analysis and for the construction and evaluation of an Affective Recognition system (Section 5.2 of the present chapter).
3. A number of overlapping windows of the raw physiological signals were separated for the feature extraction process. 743 physiological features were identified and extracted from the recorded database (Section 5.3 of the present chapter).
4. By employing a feature selection technique, a small number of the most optimal features have been identified, to reduce the dimensionality of the database to a smaller subspace, to be used in the emotion classification process (Section 5.4 of the present chapter).
5. By employing four classification algorithms (K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Discriminant Analysis (DA) and Classification Tree), the performances of a quarter of a million different classification settings (various window lengths and types, different number of features, etc.) have been evaluated and compared, using the 10-fold cross-validation technique (Sections 5.5 to 5.7 of the present chapter).

5.1. Related Research

To design the affective recognition system, in this study initially a detailed literature review over 30 affective recognition studies, published since 1993, has been conducted (refer to Table 1 in Appendix G). All aspects of these affective computing studies (e.g. stimuli type, physiological signal pre-processing, filtering, windowing, feature extraction, etc.) have been organised and reported, in detail, in seven sub-sections.

5.1.1. Stimuli and Physiological Measurements

As discussed earlier (Chapter 3), to analyse the emotional response of humans and their physiological responses, a psychophysiological affective database, recorded from a number of users exposed to a number of controlled and known affective stimuli, is required. So far various affective stimuli types have been employed in order to perform human emotional experience assessment and classification (refer to Table 2 in Appendix G). These are:

1. **Videos and Music Videos:** A considerable number of studies have concentrated on affect analysis and recognition, evoked during watching either videos[26] (Soleymani et al., 2012; Soleymani et al., 2015; Murugappan et al., 2008; Rizon et al., 2008; Bailenson et al., 2008; Koelstra & Patras, 2013; Wen et al., 2013; Yazdani et al., 2009), or music videos[27] (Soleymani et al., 2011; Koelstra et al., 2012). The average videos and music videos duration was reported as 142.76 seconds (±144.7, minimum 35 and maximum 591 seconds).
2. **Images:** The majority of the studies used images as their affective stimuli, employed extracts from the International Affective Picture System database (IAPS (Lang et al., 2008)) (Frantzidis et al., 2010; Lang et al., 1993; Jenke et al., 2014; Kukolja et al., 2014; Frantzidis et al., 2010); whilst a smaller number were found to employ other image datasets, such as affective face gestures (Othman et al., 2013). The average image presentation duration was reported as 14.5 seconds (±22.88, minimum 1 and maximum 60 seconds).
3. **Sound:** Studies, which used sound as their affective stimuli, employed either music (Takahashi & Tsukaguchi, 2003; Kim & Andre, 2008), or extracts from the International Affective Digital Sound database (IADS (Bradley & Lang, 1999)) (Nardelli et al., 2015). The average sound duration was reported as 210 seconds (±112.24, minimum 60 and maximum 330 seconds).
4. **Virtual Reality Scenarios and Games:** A number of studies employed games and virtual reality scenes as their stimuli (Wu et al., 2010; Rodríguez et al., 2015; Parnandi et al., 2013; Reuderink et al., 2013; Chanel et al., 2011; Liu et al., 2009). The average games and VRs duration was reported as 226 seconds (±106.44, minimum 120 and maximum 350 seconds).
5. **Real Life Scenario:** There are some studies, which employed real life scenarios (e.g. racing car driving) to elicit and measure various emotional experiences (Antje et al., 2005; Katsis et al., 2008; Rani et al., 2007; Rainville et al., 2006). The average stimuli duration was reported as 170 seconds (±75.5, minimum 90 and maximum 240 seconds).

Different forms of physiological measurements have been implemented in the emotional experience assessment and classification research area. The majority of studies, reviewed, measured central nervous system signals, such as EEG (Wu et al., 2010; Soleymani et al., 2012; Rodríguez et al., 2015; Frantzidis et al., 2010; Soleymani et al., 2015; Rizon et al., 2008; Murugappan et al., 2008; Koelstra & Patras, 2013; Yazdani et al., 2009; Soleymani et al., 2011; Koelstra et al., 2012; Jenke et al., 2014; Frantzidis et al., 2010; Othman et al., 2013; Takahashi & Tsukaguchi, 2003; Reuderink et al., 2013; Chanel et al., 2011; Sutton & Davidson, 2010), and

[26] Short clips from various movies scenes
[27] A short video integrating a combination of a particular song and imagery

autonomic nervous system signals, such as heart rate (either through ECG or PPG signals) (Wu et al., 2010; Lang et al., 1993; Bailenson et al., 2008; Wen et al., 2013; Soleymani et al., 2011; Koelstra et al., 2012; Kukolja et al., 2014; Takahashi & Tsukaguchi, 2003; Kim & Andre, 2008; Nardelli et al., 2015; Liu et al., 2009; Antje et al., 2005; Katsis et al., 2008; Rani et al., 2007; Rainville et al., 2006) and GSR (Wu et al., 2010; Frantzidis et al., 2010; Lang et al., 1993; Bailenson et al., 2008; Wen et al., 2013; Soleymani et al., 2011; Koelstra et al., 2012; Kukolja et al., 2014; Kim & Andre, 2008; Parnandi et al., 2013; Chanel et al., 2011; Liu et al., 2009; Antje et al., 2005; Katsis et al., 2008; Rani et al., 2007; Rainville et al., 2006) to detect affective states. There are other autonomic nervous system signals, which have been employed by a minority of studies (when compared to EEG, heart rate and GSR), such as respiratory (breathing) rate (Wu et al., 2010; Soleymani et al., 2011; Koelstra et al., 2012; Kukolja et al., 2014; Kim & Andre, 2008; Chanel et al., 2011; Katsis et al., 2008), skin temperature (Bailenson et al., 2008; Soleymani et al., 2011; Koelstra et al., 2012; Kukolja et al., 2014; Chanel et al., 2011; Liu et al., 2009; Rani et al., 2007; Antje et al., 2005), EMG (Lang et al., 1993; Soleymani et al., 2011; Koelstra et al., 2012; Takahashi & Tsukaguchi, 2003; Katsis et al., 2008; Kim & Andre, 2008; Liu et al., 2009; Rani et al., 2007; Rainville et al., 2006), blood pressure (Bailenson et al., 2008; Soleymani et al., 2011; Chanel et al., 2011) and pupil diameter (Soleymani et al., 2012; Antje et al., 2005) (refer to Table 2 in Appendix G).

5.1.2. Pre-processing (Artefact Removal and Filtering)

The signals recorded from the central nervous system are highly susceptible to noise and unwanted artefacts (Sanei & Chambers, 2009). In addition, all physiological signals recorded from the autonomic nervous system can be accompanied by unwanted noise. As an illustration, heart rate, measured using PPG, can carry unwanted noise due to harsh or rapid movements on the part of the subject's finger. Almost all studies reviewed in this paper did not apply any form of filtering on autonomic nervous system signals[28] (refer to Table 3 in Appendix G). However, to eliminate noise in EEG signals, two key techniques were found to have been employed by those studies reviewed (refer to Table 3 in Appendix G):

1. **Artefact Removal:** Muscle, eye and head movements and blinking are the main sources of noise in EEG signals (Sanei & Chambers, 2009). To minimise this type of noise, all reviewed studies instructed the participants to minimise movements as far as was possible, and to avoid excessive blinking. Half of the studies employed other artefact removal techniques, in addition, to minimise the effects of these types of noises even further. For example, (Wu et al., 2010; Rodríguez et al., 2015; Yazdani et al., 2009; Soleymani et al., 2011; Reuderink et al., 2013) employed Electrooculography (EOG) signals to attenuate eye and muscle movement. (Chanel et al., 2011; Sutton & Davidson, 2010) employed visual checking techniques, to discard portions of the EEG signal with high

[28] Only (Wen et al., 2013) applied a band-pass filter 0.5Hz to 35Hz on heart rate signal, and a low-pass filter 10Hz on GSR signal.

muscular and eye movement noises. (Frantzidis et al., 2010; Koelstra et al., 2012; Frantzidis et al., 2010) adopted other artefact removal techniques to attenuate eye and muscle movements (e.g. Least Mean Square (LMS) and Blind Source Separation (BSS) algorithms).

2. **Filtering:** Applying low, high and band-pass filters are a common technique in reducing unwanted oscillations in a signal. To minimise noises in the EEG signals, all studies applied a band-pass filter to the raw signals, with various cut-off frequencies. The majority of studies set either 4Hz or 5Hz as the lower-band cut-off frequency (Soleymani et al., 2012; Rodríguez et al., 2015; Soleymani et al., 2015; Koelstra & Patras, 2013; Soleymani et al., 2011; Koelstra et al., 2012; Takahashi & Tsukaguchi, 2003; Chanel et al., 2011), whilst others employed a variety of even lower frequencies (0.5Hz, 1Hz, etc.) (Wu et al., 2010; Frantzidis et al., 2010; Rizon et al., 2008; Murugappan et al., 2008; Yazdani et al., 2009; Jenke et al., 2014; Frantzidis et al., 2010; Reuderink et al., 2013). For the upper-band cut-off frequency, the majority of studies adopted 40Hz or 45Hz (Soleymani et al., 2012; Rodríguez et al., 2015; Frantzidis et al., 2010; Soleymani et al., 2015; Rizon et al., 2008; Koelstra & Patras, 2013; Soleymani et al., 2011; Koelstra et al., 2012; Frantzidis et al., 2010; Takahashi & Tsukaguchi, 2003; Chanel et al., 2011), whilst others employed other frequencies (30Hz, 70Hz, etc.) (Wu et al., 2010; Murugappan et al., 2008; Yazdani et al., 2009; Jenke et al., 2014; Reuderink et al., 2013; Sutton & Davidson, 2010). This was due to the fact that eye-blinking artefacts are mainly observed in frequencies lower that 4Hz, as people rarely blink more than 4 times a second. Thus selecting 4Hz as the lower-band cut-off frequency attenuates blinking effects, present in the raw EEG signals. Moreover, the brain’s high frequency rhythms (Gamma range) can be observed between 30Hz and 45Hz (Sanei & Chambers, 2009). Therefore selecting 45Hz as the upper-band cut-off frequency attenuates all higher unwanted frequencies.

5.1.3. Windowing and Physiological Feature Extraction

To perform affective recognition, a number of psychophysiological features need be extracted from the autonomic and/or central nervous system, while the participant is exposed to a number of affective stimuli (Section 5.1.1). 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. 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. (Wu et al., 2010; Rani et al., 2007; Rainville et al., 2006)).
2. The frequency analysis of physiological signals to extract specific rhythms (e.g. alpha, beta, gamma rhythm power within an EEG signal (Soleymani et al., 2012; Kim & Andre, 2008)).

3. The detection of specific patterns, such as Event Related Potentials (ERP – such as the P300, N100, and others (Frantzidis et al., 2010; Frantzidis et al., 2010; Yazdani et al., 2009)).
4. Other exclusive measurements (e.g. EEG_w[35] (Chanel et al., 2011)).

Table 16 presents a list of affective features, employed by the reviewed studies, which have been reported to be related to emotional states. Only the features extracted from the most commonly used physiological signals (EEG, GSR and Heart Rate – as reported in Section 5.1.1) are covered here (refer to Table 5 to 7 in Appendix G).

Table 16 – Affective Features Related to Emotional States, Reported in the Reviewed Literature

Feature	Used in Studies	Feature	Used in Studies
EEG Single Channel Theta Rhythm	(Wu et al., 2010), (Soleymani et al., 2012), (Rodríguez et al., 2015), (Soleymani et al., 2015), (Koelstra & Patras, 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Jenke et al., 2014), (Chanel et al., 2011)	GSR Minimum	(Wu et al., 2010)
EEG Single Channel Slow Alpha Rhythm	(Soleymani et al., 2012), (Soleymani et al., 2015), (Koelstra & Patras, 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Jenke et al., 2014)	GSR Maximum	(Wu et al., 2010), (Wen et al., 2013), (Liu et al., 2009), (Rani et al., 2007)
EEG Single Channel Alpha Rhythm	(Wu et al., 2010), (Soleymani et al., 2012), (Rodríguez et al., 2015), (Soleymani et al., 2015), (Koelstra & Patras, 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Jenke et al., 2014), (Reuderink et al., 2013), (Chanel et al., 2011), (Sutton & Davidson, 2010)	GSR Standard Deviation	(Wen et al., 2013), (Kim & Andre, 2008)

EEG Single Channel Beta Rhythm	(Wu et al., 2010), (Soleymani et al., 2012), (Soleymani et al., 2015), (Koelstra & Patras, 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Jenke et al., 2014), (Chanel et al., 2011)	GSR Mean of the First Derivative	(Wen et al., 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Kim & Andre, 2008), (Chanel et al., 2011), (Liu et al., 2009), (Katsis et al., 2008), (Rani et al., 2007)
EEG Single Channel Gamma Rhythm	(Soleymani et al., 2012), (Soleymani et al., 2015), (Koelstra & Patras, 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Jenke et al., 2014)	GSR Mean of the Negative Values in the First Derivative	(Soleymani et al., 2011), (Koelstra et al., 2012), (Chanel et al., 2011)
EEG Paired Channel[29] Theta Rhythm	(Soleymani et al., 2012), (Koelstra & Patras, 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Jenke et al., 2014)	GSR Mean of the Peak Values	(Soleymani et al., 2011), (Koelstra et al., 2012)
EEG Paired Channel[29] Slow Alpha Rhythm	(Koelstra & Patras, 2013), (Jenke et al., 2014)	GSR Low Frequency[30] Spectral Power	(Soleymani et al., 2011), (Koelstra et al., 2012)
EEG Paired Channel[29] Alpha Rhythm	(Soleymani et al., 2012), (Koelstra & Patras, 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Jenke et al., 2014)	Heart Rate Mean	(Wu et al., 2010), (Bailenson et al., 2008), (Wen et al., 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Takahashi & Tsukaguchi, 2003), (Kim & Andre, 2008), (Nardelli et al., 2015), (Liu et al., 2009), (Antje et al., 2005), (Katsis et al., 2008), (Rani et al., 2007), (Rainville et al., 2006)
EEG Paired Channel[29] Beta Rhythm	(Soleymani et al., 2012), (Koelstra & Patras, 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Jenke et al., 2014)	Heart Rate Standard Deviation	(Wen et al., 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Kukolja et al., 2014), (Nardelli et al., 2015), (Rani et al., 2007), (Rainville et al., 2006)

[29] Subtraction of the two channels' raw signals
[30] GSR low frequency is 0Hz to 2.4Hz

EEG Paired Channel[29] Gamma Rhythm	(Soleymani et al., 2012), (Koelstra & Patras, 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Jenke et al., 2014)	Heart Rate Mean of the Peak Values	(Liu et al., 2009)
Asymmetric Spectral Power Density[31] Using Alpha Rhythm	(Rizon et al., 2008), (Murugappan et al., 2008)	Heart Rate Mean of the First Derivative	(Wen et al., 2013), (Kim & Andre, 2008)
Asymmetric Spectral Variance Density[32] Using Alpha Rhythm	(Rizon et al., 2008), (Murugappan et al., 2008)	Heart Rate Low Frequency[33] Spectral Power	(Bailenson et al., 2008), (Soleymani et al., 2011), (Koelstra et al., 2012), (Kukolja et al., 2014), (Kim & Andre, 2008), (Nardelli et al., 2015), (Rainville et al., 2006)
Event Related Potentials (EPR)	(Frantzidis et al., 2010), (Yazdani et al., 2009), (Frantzidis et al., 2010)	Heart Rate Medium Frequency[34] Spectral Power	(Bailenson et al., 2008), (Soleymani et al., 2011), (Koelstra et al., 2012), (Kukolja et al., 2014), (Kim & Andre, 2008), (Nardelli et al., 2015), (Rainville et al., 2006)
EEGw[35]	(Chanel et al., 2011)	Heart Rate High Frequency[36] Spectral Power	(Bailenson et al., 2008), (Soleymani et al., 2011), (Koelstra et al., 2012), (Kukolja et al., 2014), (Kim & Andre, 2008), (Nardelli et al., 2015), (Rainville et al., 2006)
GSR Mean	(Wu et al., 2010), (Bailenson et al., 2008), (Wen et al., 2013), (Soleymani et al., 2011), (Koelstra et al., 2012), (Kukolja et al., 2014), (Kim & Andre, 2008), (Chanel et al., 2011), (Liu et al., 2009), (Antje et al., 2005), (Katsis et al., 2008), (Rani et al., 2007), (Rainville et al., 2006)	Heart Rate Spectral Power Ratio[37]	(Soleymani et al., 2011), (Koelstra et al., 2012), (Kukolja et al., 2014), (Nardelli et al., 2015)

[31] Asymmetric Spectral Power Density $= \frac{P_{left}-P_{right}}{P_{left}+P_{right}}$, While "P" is the spectral power in "Alpha" frequency rhythms.

[32] Asymmetric Spectral Variance Density $= \frac{V_{left}-V_{right}}{V_{left}+V_{right}}$, While "V" is the spectral power in "Alpha" frequency rhythms.

[33] Heart Rate low frequency is 0.01Hz to 0.04Hz.

[34] Heart Rate Medium frequency is 0.04Hz to 0.15Hz.

[35] $EEG_w = \log\left(\frac{\sum_{i=1}^{N}\beta_i}{\sum_{i=1}^{N}(\theta_i+\alpha_i)}\right)$, While "N" is the number of channels. θ, α and β are Theta, Alpha and Beta frequency rhythms.

[36] Heart Rate High frequency is 0.15Hz to 0.5Hz.

[37] Heart Rate Spectral Power Ration $= \frac{\text{Medium Frequency Spectral Power}}{\text{High Frequency Spectral Power}}$

To extract the affective features (employing statistical, spectral, pattern or/and other measurements), 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. Three main algorithms have been employed in the literature, to extract the appropriate periods of the physiological signals, which can be confidently tagged by a specific affective state (refer to Table 4 in Appendix G):

1. **Window with Entire Stimuli Length:** 46% of the reviewed studies extracted the features from the entire duration of the stimuli, regardless of the length. As an illustration, if participants were exposed to affective images for 5 seconds, the affective features were extracted from the 5-second affective period. Similarly, if participants were exposed to affective music videos for 60 seconds, the affective features were extracted from the 60-second affective period (Rodríguez et al., 2015; Rizon et al., 2008; Murugappan et al., 2008; Bailenson et al., 2008; Koelstra & Patras, 2013; Jenke et al., 2014; Kukolja et al., 2014; Othman et al., 2013; Takahashi & Tsukaguchi, 2003; Reuderink et al., 2013; Liu et al., 2009; Rainville et al., 2006; Sutton & Davidson, 2010).
2. **Window Shorter than Stimuli Length:** 36% of the reviewed studies extracted the features from multiple windows within each stimulus (i.e. shorter than the overall stimuli duration). The majority of these studies employed non-overlapped windowing techniques (Wu et al., 2010; Wen et al., 2013; Yazdani et al., 2009; Kim & Andre, 2008; Nardelli et al., 2015; Parnandi et al., 2013; Chanel et al., 2011; Katsis et al., 2008), and minority have used windows with 50% overlap (Soleymani et al., 2012; Soleymani et al., 2015).
3. **Window Longer than Stimuli Length:** Only 18% of the studies included the post or/and pre-stimuli periods, and extracted the features from windows longer than the stimuli length (Frantzidis et al., 2010; Lang et al., 1993; Soleymani et al., 2011; Koelstra et al., 2012; Frantzidis et al., 2010).

5.1.4. Normalisation

Physiological signals can, of course, be subject-sensitive. As an illustration, different skull thickness could cause various EEG voltage levels (Lehtinen et al., 1996). To compensate for these types of variations, 43% of reviewed studies employed normalisation techniques, on either **extracted features** or **raw signals**, to enable proper comparison. However, 57% of the reviewed literature did not apply any normalisation techniques, in their study. So far, it has been found that researchers, in the main, have employed three different normalisation techniques (refer to Table 2 in Appendix G). Consider $X_R(t)$ as a signal recorded over time:

1. **Min-Max Normalisation:** The min-max normalised $X_N(t)$ can be calculated using Equation 3 ("min" and "max" represent the minimum and maximum

values of $X_R(t)$, respectively, over time). The extreme values of the normalised signal would be between 0 and 1. (Wu et al., 2010; Soleymani et al., 2012) employed this technique to normalise the **extracted features** across all participants' observations, while (Yazdani et al., 2009; Liu et al., 2009) used this algorithm to normalise the **raw recorded signals** within each participant's observation (either within each window or across the entire stimuli duration).

$$X_N(t) = \frac{X_R(t) - \min(X_R(t))}{\max(X_R(t)) - \min(X_R(t))}$$

Equation 3 – Min-Max Normalisation Equation

2. **Z-Score Normalisation:** The z-score normalised $X_N(t)$ can be calculated using Equation 4 ("mean" and "std", represent the average and standard deviation of $X_R(t)$, respectively, over time). The extreme values of the normalised signal would be between about -3 and 3 (Kreyszig et al., 2010). (Soleymani et al., 2015; Takahashi & Tsukaguchi, 2003; Jenke et al., 2014) employed this technique to normalise the **extracted features** across all participants' observations, while (Rizon et al., 2008; Murugappan et al., 2008; Wen et al., 2013; Kim & Andre, 2008) used this algorithm to normalise the **raw recorded signals** within each participant's observation (either within each window or across the entire stimuli).

$$X_N(t) = \frac{X_R(t) - \mathrm{mean}(X_R(t))}{\mathrm{std}(X_R(t))}$$

Equation 4 – Z-Score Normalisation Equation

3. **Log-Transformation:** The log-transformed $X_N(t)$ can be calculated using Equation 5 (α and β are two constant arbitrary values). The log-transformation makes highly skewed distributions less skewed, as any 1% changes in $X_R(t)$ values would cause an average of $\frac{\beta}{100}$ unit changes in $X_N(t)$ (Lane et al., 2013). (Sutton & Davidson, 2010) employed this technique to normalise the **extracted features** across all participants' observations, while (Lang et al., 1993) used this algorithm to normalise the **raw recorded signals** within each participant's observation (either within each window or across the entire stimuli).

$$X_N(t) = \alpha + \beta log_{10}(X_R(t))$$

Equation 5 – Log-Transformation Equation

5.1.5. Affective Analysis, Classification and Validation

Only 13% of the reviewed papers investigated the significance of extracted affective features' variations with respect to the emotional states (using t-Test, Analysis of Variance (ANOVA), correlation, etc.) (Othman et al., 2013; Parnandi et al., 2013; Reuderink et al., 2013; Antje et al., 2005); while the remaining papers performed affective recognition. 13% of the reviewed papers employed regression algorithms to design continuous emotion detection systems (Lang et al., 1993; Soleymani et al., 2015; Soleymani et al., 2011; Liu et al., 2009). The other 74%

designed and trained classifiers to detect emotional responses, according to a number of defined affective clusters. Table 17 presents the different types of classifiers employed in the reviewed studies. On average, studies recruited 25.1 participants (±20.6, with a minimum of 3 and a maximum of 101), to train and validate their affective recognition system (refer to Table 2 in Appendix G).

Table 17 – Classifiers Employed in Studies

Classifier	Used in Studies	Classifier	Used in Studies
Support Vector Machine (SVM)	(Wu et al., 2010), (Soleymani et al., 2012), (Frantzidis et al., 2010), (Takahashi & Tsukaguchi, 2003), (Chanel et al., 2011), (Liu et al., 2009), (Katsis et al., 2008)	Neural Network (NN)	(Bailenson et al., 2008), (Takahashi & Tsukaguchi, 2003)
Discriminant Analysis (DA)	(Yazdani et al., 2009), (Jenke et al., 2014), (Kim & Andre, 2008), (Chanel et al., 2011), (Rainville et al., 2006)	Gaussian Naïve Bayes (GNB)	(Koelstra & Patras, 2013), (Koelstra et al., 2012)
Classification Tree	(Wu et al., 2010), (Frantzidis et al., 2010), (Liu et al., 2009), (Rani et al., 2007)	Random Forest	(Wen et al., 2013)
K^{th} Nearest Neighbour (KNN)	(Kukolja et al., 2014), (Liu et al., 2009)	Quadratic Bayes Normal (QBN)	(Nardelli et al., 2015)
Fuzzy C-Mean	(Rizon et al., 2008), (Murugappan et al., 2008)	Fuzzy Logic	(Rani et al., 2007)

5.1.6. Classification/Regression Performance Evaluation

Cross-validation techniques are considered to be one of the most popular algorithms in classification performance evaluation processes. This is due to the fact that in the majority of cases, the entire recorded database is employed as both the training and test dataset. All studies, which were found to perform either classification or regression, have been subjected to cross-validations to assess the performance of their affective recognition systems. However, there are 3 different types of cross-validation techniques, which are described in the literature:

1. **K-Fold:** In this technique, the dataset is divided into *K* randomly selected folds[38]. Then *K* iterations are executed to train a classifier using *K-1* folds of the data and to evaluate its performance on the remaining fold. The performance of the classifier is reported according to the average performances of all *K*

[38] If a dataset is divided into a number of symmetrical portions (subsamples), each portion is called a fold.

iterations (Frantzidis et al., 2010; Chanel et al., 2011; Koelstra & Patras, 2013; Soleymani et al., 2015).

2. **Leave-One-Observation/Session-Out:** In this method, one single observation or session is taken out of the dataset. Then a classifier is trained using the remaining data and tested on the subtracted single observation or session. The performance of the classifier is reported according to the average performances of all iterations (equal to the number of observations or sessions) (Takahashi & Tsukaguchi, 2003; Rainville et al., 2006; Rani et al., 2007; Katsis et al., 2008; Kim & Andre, 2008; Liu et al., 2009; Soleymani et al., 2011; Parnandi et al., 2013; Jenke et al., 2014; Kukolja et al., 2014).
3. **Leave-One-Participant-Out:** In this approach, the entire observations, generated by a single participant, are taken out of the dataset. Then a classifier is trained using the remaining data and tested on the subtracted participant. The performance of the classifier is reported according to the average performances of all iterations (equal to the number of participants) (Bailenson et al., 2008; Wu et al., 2010; Soleymani et al., 2012; Koelstra et al., 2012; Wen et al., 2013; Nardelli et al., 2015).

The studies, which implemented classification algorithms for emotion recognition processes (Table 17), reached an average of 77.89% classification accuracy (±11.53%, with minimum of 59.7% and maximum of 96.5%), calculated through cross-validation techniques (refer to Table 2 in Appendix G).

5.1.7. Emotion Assessment

As discussed in Chapter 2, two distinct affective models could be employed to perform emotion assessments; dimensional and categorical. Almost 40% of the affect analysis and recognition papers reviewed in this study, employed a categorical model of emotion to perform emotional assessments (Rodríguez et al., 2015; Murugappan et al., 2008; Rizon et al., 2008; Bailenson et al., 2008; Wen et al., 2013; Yazdani et al., 2009; Kukolja et al., 2014; Chanel et al., 2011; Liu et al., 2009; Katsis et al., 2008; Rani et al., 2007; Rainville et al., 2006); whereas the other 60% employed the dimensional model of affect to perform emotion assessments (Wu et al., 2010; Soleymani et al., 2012; Frantzidis et al., 2010; Lang et al., 1993; Soleymani et al., 2015; Koelstra & Patras, 2013; Soleymani et al., 2011; Koelstra et al., 2012; Jenke et al., 2014; Frantzidis et al., 2010; Othman et al., 2013; Takahashi & Tsukaguchi, 2003; Kim & Andre, 2008; Nardelli et al., 2015; Parnandi et al., 2013; Reuderink et al., 2013; Antje et al., 2005). Moreover, as discussed in Section 2.2, to perform the emotional experience assessment (regardless of the employed model of affect), three different techniques have been employed by the studies. 13% of the studies employed pre-affective hypothesis (Bailenson et al., 2008; Othman et al., 2013; Takahashi & Tsukaguchi, 2003; Parnandi et al., 2013) and only one study employed expert assessment (Katsis et al., 2008). However, majority of the studies (83%) employed self-assessment techniques (Frantzidis et al., 2010; Jenke et al., 2014; Frantzidis et al., 2010; Nardelli et al., 2015; Soleymani et al., 2011; Soleymani et al., 2012; Soleymani et al., 2015; Rodríguez et al., 2015; Lang et al., 1993; Rizon et al., 2008; Murugappan et al., 2008; Koelstra & Patras, 2013; Wen et al., 2013; Yazdani et al., 2009; Koelstra et al., 2012; Kukolja et al., 2014; Reuderink et al., 2013; Chanel et al., 2011; Liu et al., 2009; Antje et al., 2005; Rani et al., 2007; Rainville et al., 2006; Sutton & Davidson, 2010) (refer to Table 2 in Appendix G).

5.2. Psychophysiological Database Construction (Primary Experiment – *Experiment 4*)

5.2.1. Material

5.2.1.1. Affective Virtual Reality

In the present study, the two most affective games, in each of the four Affective Clusters[39] introduced in Section 2.3.5, have been identified using the Cosine Similarity Algorithm (Pang-Ning Tan, 2005) as implemented in Section 3.5.2 (refer to Appendix D). As a result of this analysis, the eight most 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 affective games, two '*neutral games*' were added in the experiment (the neutral game in Section 3.5.2, 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 the present experiment.

5.2.1.2. Participant Selection

One of the most important challenges of designing any affective psychophysiological database is the minimisation of variability between participants, in each individual affective session, whilst maximising the variability between sessions' experiences. This is due to the fact that, in any human-cantered experiment, minimum variability between participants' experiences, in a single affective session is an extremely important issue. Any acceptable analysis, dealing with either affects or physiological databases, should, intuitively, be based on changes in emotional experiences, due to **different environments**, rather than different **personal experiences**. As discusses in Section 4.3.3, the Multi-variant Analysis of variance (MANOVA) highlighted significant differences between the four participant groups (male gamers, male non-gamers, female gamers and female non-gamers). The analysis showed that male gamers, male non-gamers and female gamers show marked similarities in their affective experiences, when compared to female non-gamers. Therefore, in the present study, in order to minimise *between participants* variability, it was decided to recruit only **male and female gamers** in the experiment. Recruiting the male non-gamers (despite their highest level of similarity compared to the others) in the experiment would disturb the comparison, as no female non-gamers were to be recruited. Moreover, it was highlighted that the participants, aged between 18 and 30 had more similar emotional responses, when compared to those who were 30 to 40

[39] (1) Positive Valence and Low Arousal (PVLA) – (2) Positive Valence and High Positive Arousal (PVHPA) – (3) Negative Valence and Positive Arousal (NVPA) – (4) Negative Valence and Negative Arousal (NVNA) – Table 3.

years old. Therefore, 30 gamer participants of both genders (15 of each), aged between 18 and 30 years old (mean age=22.76), were recruited to take part in this experiment. Each participant received a £10 gift voucher at the end of each experiment. The study was reviewed and approved by the University of Birmingham's Ethical Review Committee (Ethical Reference Number: ERN_13-1157).

5.2.1.3. Physiological Signal Recording

As discussed in Section 5.1.1, the majority of studies have employed EEG, Heart Rate and GSR signals to perform affective analysis and recognition. Therefore, 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 EMOTIV EPOC[40] headset to record EEG signals, as well as Shimmer+ wearable sensor technologies[41] to record GSR and heart rate activities. The EMOTIV EPOC records the EEG signals, with a 128Hz sampling frequency, from 14 channels[42]. The electrodes are 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 512Hz sampling frequency (refer to Appendix H).

A program was developed to function in parallel with the game, to establish a connection with the Shimmer+ and EMOTIV EPOC devices (through the software development tool kits (SDKs) provided by the manufacturers), as soon as each game was started. As well as signal recording, the program performed time synchronisation, to align the outputs received from the devices, prior to data storage. Furthermore, a wireless real-time monitoring application (tablet-based) was implemented, to enable the experimental supervisor to monitor the software functionality and signal qualities of the recording devices, without any distraction to the participants' experiences (refer to Appendix H).

5.2.2. 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 Section 3.5.3). 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 both dimensional (Valence, Arousal and Dominance) and categorical (according to eight Emotion Labels: Relaxed, Content,

[40] http://emotiv.com/
[41] http://www.shimmersensing.com/
[42] 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.

Happy, Excited, Angry, Afraid, Sad and Bored) models of affect (as presented in Section 2.3.1). The participants were given a 5- to 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 (refer to Appendix E and I for the "Consent Form" and "Information Sheet").

5.2.3. Results (Psychophysiological Database)

Of a possible total of 300 affective sessions, 290 were recorded, as 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 EMOTIV EPOC headset[43] 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. The Shimmer+ software uses the estimation techniques introduced in (K. Nakajima, 1996) to approximate the heart rate of the participants 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 (refer to Appendix H). The emotional ratings of the participants were recorded and stored separately at the end of each game.

5.3. Feature Matrix Construction

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

5.3.1. Pre-Processing

5.3.1.1. Filtering

As discussed in Section 5.1.2, the majority of affective recognition studies applied no filtering technique to heart rate and GSR signals recorded from the participants. Therefore in this study, we also decided to use the heart rate and GSR signals without applying any filter. However, and as discussed in Section 5.1.2, the majority of previous studies employed a band-pass filter (majority using 4Hz to 45Hz) in their EEG filtering process. Eye-blink artefacts are typically observed in frequencies lower than 4Hz, as humans rarely blink more than 4 times a second. In addition, high-frequency rhythms in the brain (Gamma range) can be observed from 30Hz up to 45Hz (Sanei & Chambers, 2009). Therefore, in the present study, a 5th order Butterworth band-pass filter was applied to the raw EEG signals, whilst the lower-band was set to 4Hz and the upper-band was fixed at 45Hz.

As discussed in Section 5.1.2, the majority of artefact removal techniques were performed either by the use of Electrooculography (EOG) signals, or by using computationally expensive EEG artefact removal algorithms, such as those based on

[43] According to five classes; good, fair, poor, very bad and disconnected.

Least Mean Square (LMS) and Blind Source Separation (BSS). Due to the absence of EOG signal recording in this study, and the high computational expense of other artefact removal algorithms, it was decided to apply **NO** EEG artefact removal technique.

5.3.1.2. Normalisation

As discussed in Section 5.1.4, almost half of the reviewed studies employed a normalisation technique. However, half of the studies, which employed normalisation techniques, normalised the recorded **raw signals**, while the rest normalised the **extracted features**. Normalisation can improve the accuracy of regression or classification techniques, as it can minimise the between-participants differences, which could be present in any physiological measurements (e.g. average heart rate can vary between people, different skull thicknesses could change the power of the EEG signals (Lehtinen et al., 1996), etc.). However, any normalisation technique needs to be calibrated according to a recorded dataset [44] (Section 5.1.4). Implementation of normalisation techniques can be considered as an advantage for offline classifiers, as the required features or raw signals would be extracted from the pre-recorded raw signals, before the classification process. This can provide all of the required data for the normalisation technique for the calibration process. Whereas in online classifiers the required features need to be extracted from the progressing raw signals, while the classification is in progress. This issue can turn the normalisation leverage to a disadvantage, as the normalisation calibration process could perform inadequately in the absence of the entire required dataset.

In the present study, we decided to avoid feature normalisation techniques. However, before performing a spectral analysis on the raw signals (Fast Fourier Transform (FFT) – refer to Section 5.3.3.1), we applied the ***z*-score normalisation technique** (Section 5.1.4) within EEG channels, to standardise the spectral analysis. As an illustration, the thickness of one part of a participants' head may vary from another, or the channels' signal contact quality[45] may slightly vary. These small changes could result in considerable variation between the signal powers of the EEG channels (Lehtinen et al., 1996). Therefore, applying normalisation on the raw signals can standardize the power spectrum comparison between channels, as the overall power of each normalised channel could be between around -3 and 3 (Section 5.1.4), regardless of the corresponding skull thickness and signal contact quality.

5.3.2. Windowing

As discussed in Section 5.1.3, different affective stimuli with various durations have been employed in the studies. Therefore it has not been possible to use the available literature to define the most appropriate window length, with respect to the affective stimuli duration.

[44] Such as minimum or maximum values of the signal in min-max normalisation technique, α and β in log-transformation algorithm (Equation 5), etc.

[45] The signal quality of the channels depends on the wetness of the EEG electrodes foams. Although the appropriate preparations have been conducted, at the beginning of each experiment, to ensure proper channels' signal quality; it cannot be stated, with confidence, that all channels are recorded with exactly equal contact quality (i.e. exactly equal wetness).

5.3.2.1. Window Length

As the majority of the studies have employed window lengths that are either equal to, or shorter than the stimuli duration (Section 5.1.3), we decided to disregard the post- or pre-stimuli duration and, thus, to avoid any window length longer than the stimuli durations. To assess the efficiency of different window lengths, we selected 28 arbitrary durations according to two different algorithms:

1. **Fixed Duration:** In this method, the duration of windows would be a **fixed value** in all sessions, for all participants, regardless of the stimuli duration. In this method we arbitrarily selected the following (17) fixed window durations: 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 100, 150 and 200 seconds (if applicable; i.e. the window length is not longer than the game duration).
2. **Relative Duration:** In this method, the duration of the windows would be calculated in every session, independently, according to the session duration. To do so, a global **relative value** was selected (as a percentage of the stimuli duration), to enable the system to calculate the window length according to the duration of the stimuli. In this method we arbitrarily selected the following (11) relative window durations: 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 60%, 80% and 100%.

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.

5.3.2.2. Window Type

To perform spectral analysis on the signals, we employed a Fast Fourier Transform (FFT) technique (Section 5.3.3.1). One of the hypotheses of the FFT analysis technique is the periodicity of the target signal (Press et al., 1992). 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 (Harris, 1978). To eliminate this effect, **weighting window functions** can be applied to the signal before FFT analysis takes place (Harris, 1978). If the weighting window function is made of N elements, there are N coefficient weights (called W vector) for all corresponding elements within the window. In the present study, two weighting window functions are implemented and evaluated:

1. **Hamming Window:** This window multiplies the values closer to the centre of the window by a coefficient close to one, and the values closer to the edges by a weight closer to zero. Equation 6 and Figure 19 present the Hamming window coefficients formula, and shape, respectively (Harris, 1978).

$$W(n) = 0.54 + 0.46\, cos\left(2\pi \frac{n}{N}\right), 0 \leq n \leq N$$

Equation 6 – Hamming Window Coefficient Formula

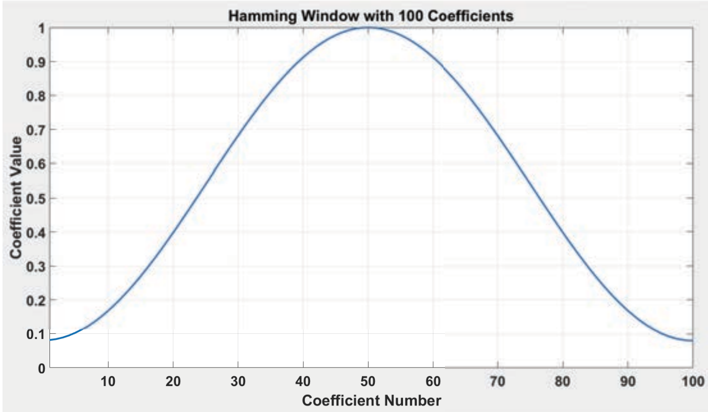


Figure 19 – Hamming Window With 100 Coefficients

2. **Tukey Window:** This window is similar to a rounded-edge rectangular window, with a parameter r, which can be tuned between 0 and N. Equation 7 and Figure 20 present the Tukey window coefficients formula, and shape, respectively (Harris, 1978). In the present study, we fixed the value of r at $\frac{N}{2}$.

$$W(n) = \begin{cases} 0.5\left(1 + cos\left(\frac{2\pi}{r}\left[n - \frac{r}{2}\right]\right)\right), & 0 \leq n \leq \frac{r}{2} \\ 1, & \frac{r}{2} \leq n \leq N - \frac{r}{2} \\ 0.5\left(1 + cos\left(\frac{2\pi}{r}\left[n - N + \frac{r}{2}\right]\right)\right), & N - \frac{r}{2} \leq n \leq N \end{cases}$$

Equation 7 – Tukey Window Coefficient Calculation Formula

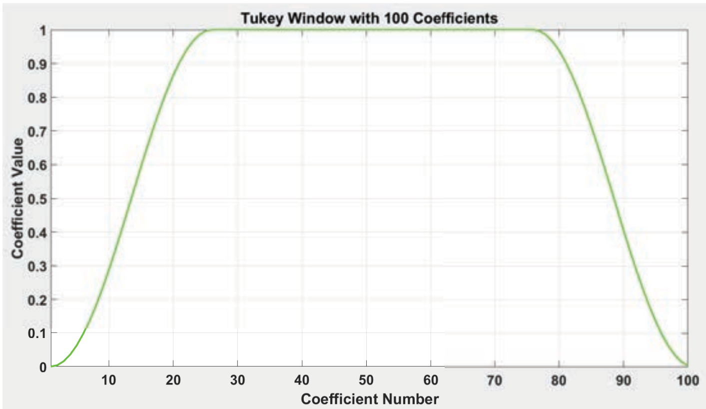


Figure 20 – Tukey Window With 100 Coefficients, While $r = \frac{N}{2} = 50$

5.3.2.3. Window Overlap

As explained above (Section 5.3.2.2), the signal values are attenuated, due to the window coefficient weights, before spectral analysis[46]. Therefore, by applying non-overlapped windows, almost 50% and 30% of the signal values, passed through Hamming and Tukey windows, respectively, 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. Figure 21 presents five Hamming windows, with 50% overlap, with 100 samples, over 200 data points. The overlapping areas are the most attenuated data points in the entire windowed signal, as they are attenuated on both windows. In the present study it was decided to avoid any maximum attenuation, larger than (around) 5%, at all signal points. To achieve that, the Hamming window was shifted every $\frac{N}{6}$ samples, to create 83.34% overlap, with about 5.5% maximum attenuation; and the Tukey window was shifted every $\frac{N}{2}$ samples, to create 50% overlap, with about 0.5% maximum attenuation.

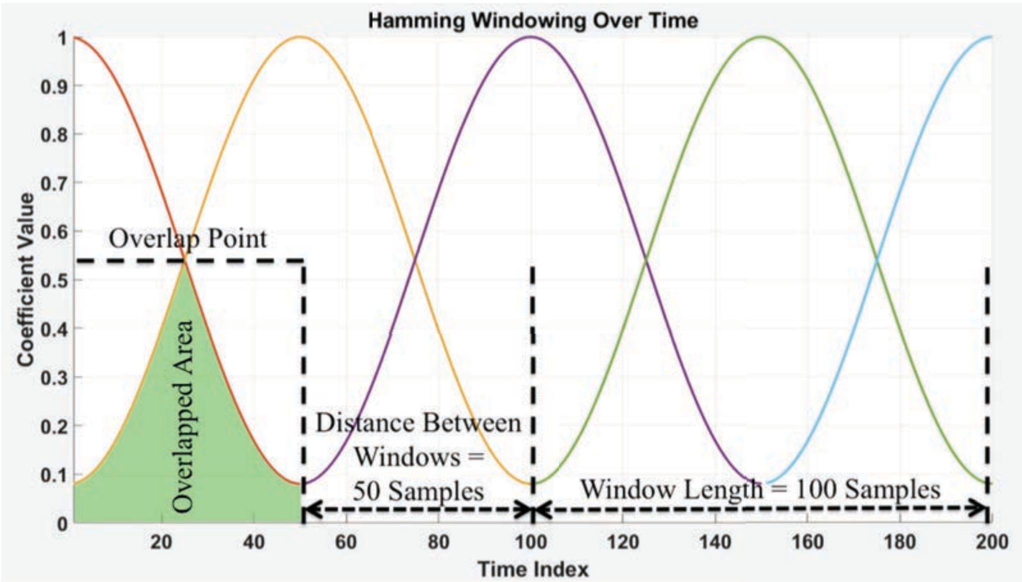


Figure 21 – 50% Overlapped, 100-Sample Hamming Windows, Over a 200-Sample Signal

5.3.2.4. Windowing Parameters

As discussed in Sections 5.3.2.1, 5.3.2.2 and 5.3.2.3, there are two parameters in the windowing process; (1) **Duration** and (2) **Type**. For duration, there are 28 arbitrary choices, which have been implemented in this study; (I) fixed (we used 17 arbitrary lengths) and (II) relative (we used 11 arbitrary values). There are two window types that have been used in this study; (I) Hamming and (II) Tukey

[46] 50% of the signal values in Tukey window, with $r = \frac{N}{2}$, are attenuated with values smaller than 1, and only 1 point (window centre value) in the Hamming window would not be attenuated.

windows. Therefore, the combinations can create 56 different windowing processes[47]. The optimum selection of these parameters is investigated in Section 5.6.2.

5.3.3. Training Features Matrix

The training features matrix is an n by m matrix, where n represents the number of observations (windows) and m signifies the number of features. Each row represents m features ($F_i = [f_{i1} \quad \cdots \quad f_{im}]_{1\times m}$, i^{th} observation), extracted from a single window, with observed output, y_i. Equation 8 presents the relationship between the training feature matrix and the predicted outputs. Function g (classifier, regression function, etc.) predicts $\hat{y}_i$, at any given point, given the corresponding features matrix F_i (Murphy, 2012, pp.2-12). To construct the training affective features matrix, four steps have been followed in the present study:

1. The essential affective features, for all physiological measurements have been identified and extracted, for each window (F_i - in total 743 features per window – Sections 5.3.3.3, 5.3.3.4 and 5.3.3.5).
2. Each row of the features matrix is time-stamped with its corresponding window centre time.
3. Each row of the features matrix is tagged by the affective rating, self-reported by the participant at the end of each game (Section 5.3.3.6).
4. The defective rows of the features matrix have been identified and deleted from the final training features matrix (Section 5.3.3.7).

$$g\left\{\begin{bmatrix} f_{11} & \cdots & f_{1m} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nm} \end{bmatrix}_{n\times m}\right\} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}_{n\times 1}$$

Equation 8 – Training Features matrix Vs. Outputs Function

5.3.3.1. Spectral Analysis

In this study, the Fast Fourier Transform (FFT) technique (Press et al., 1992) was employed, for the extraction process of all spectral features. To extract any given frequency bandwidth power, four calculation techniques have been implemented. In all 4 equations, $Power_{fr}$ indicates the FFT power of a specific frequency (fr).

1. **Power Summation:** In this technique, the power of all frequency samples within the required frequency bandwidth (from frequency fr_1 to fr_2), are added, to generate the overall power summation (Equation 9). This measurement derives the simple accumulative power, within a particular frequency bandwidth.

$$Power\ Summation = \sum_{fr=fr_1}^{fr_2} Power_{fr}$$

Equation 9 – Power Summation Equation

[47] $2\times(17+11)=56$

2. **Power Ratio:** In this technique the squared power of all frequency samples within the required bandwidth (from frequency fr_1 to fr_2) are added, and then divided by the accumulated squared power of all frequency bandwidths in the signal (from frequency fr_{min} to fr_{max}), to generate the overall power ratio (Equation 10). This measurement derives a fractional power unit, within a particular frequency bandwidth, with respect to all other bandwidths.

$$Power\ Ratio = \frac{\sum_{fr=fr_1}^{fr_2}(Power_{fr})^2}{\sum_{fr=fr_{min}}^{fr_{max}}(Power_{fr})^2}$$

Equation 10 – Power Ratio Equation

3. **RMS Power:** In this technique the Root Mean Square (RMS) power of all frequency samples within the required bandwidth (from frequency fr_1 to fr_2) is calculated (Equation 11). N is the number of frequency samples, available within the corresponding bandwidth.

$$RMS\ Power = \sqrt{\frac{\sum_{fr=fr_1}^{fr_2}(Power_{fr})^2}{N}}$$

Equation 11 – RMS Power Equation

4. **RMS Power Ratio (db):** In this technique, the logarithmic measure of the Root Mean Square (RMS) power is calculated. To do so, the RMS power of the required frequency bandwidth (from frequency fr_1 to fr_2) is calculated. Then, the logarithmic measure of this value, divided by the RMS power of all frequency bandwidths (from frequency fr_{min} to fr_{max}), is calculated. N and M are the number of frequency samples, available within the corresponding bandwidth and all bandwidths, respectively (Equation 12). The unit in this measurement is the decibel (db). This measurement derives a normalised power unit of the RMS power of a frequency bandwidth, respect to the RMS power of all frequency bandwidths[48].

$$RMS\ Power\ Ratio = 10log_{10}\left(\frac{\sqrt{\frac{\sum_{fr=fr_1}^{fr_2}(Power_{fr})^2}{N}}}{\sqrt{\frac{\sum_{fr=fr_{min}}^{fr_{max}}(Power_{fr})^2}{M}}}\right)$$

Equation 12 – RMS Power Ratio (db) Equation

[48] The Log-Transformation, discussed in Section 5.1.4, while $\beta = 10$ and $\alpha = -10log_{10}\left(\sqrt{\frac{\sum_{fr=fr_{min}}^{fr_{max}}(Power_{fr})^2}{M}}\right)$ as a reference point.

5.3.3.2. Participant-Related 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 (refer to Appendix J for features indices and the corresponding values specifications – Features Indices 2 to 4).

5.3.3.3. EEG Features

Table 18 presents all features extracted from the EEG signals. 13 out of 14 EEG features, presented in Section 5.1.3, have been extracted from the EEG signal within each window. Only the event-related potentials (EPR) have not been employed in this study. This was due to the fact that the affective stimulations were presented during the game period, whereas EPR features need to be extracted according to specific stimuli presentation instances. In addition to the 13 EEG features introduced in Section 5.1.3, the $Alpha - Beta\ Ratio$ measurement, presented in Equation 13, has been implemented in this study. According to (Sanei & Chambers, 2009), 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 (large $Alpha - Beta\ Ratio$ indicates high alpha activities and lower beta activations, signifying lower attention and concentration, and vice versa).

$$Alpha - Beta\ Ratio = \frac{Alpha\ Bandwidth\ Power}{Beta\ Bandwidth\ Power}$$

Equation 13 – Alpha-Beta Ratio Equation

All spectral powers, in all frequency ranges (theta, slow-alpha, alpha, beta and gamma), have been extracted four times, using one of the power calculation formulae presented in Section 5.3.3.1 (Equation 9, Equation 10, Equation 11 and Equation 12). Therefore all related measurements (e.g. asymmetric ratio[49], EEG_w[50], etc.) have been calculated four times, using all four power calculation formulas. Consequently, in total, 707 features have been extracted from all fourteen single and seven paired channels of EEG signals (refer to Appendix J for features indices and the corresponding values specifications – Features Indices 20 to 726).

[49] Asymmetric Spectral Power Density $= \frac{P_{left} - P_{right}}{P_{left} + P_{right}}$, While "P" is the spectral power in either "Alpha" or "Slow-Alpha" frequency rhythms (Murugappan et al., 2008), (FRANTZIDIS, Christos A. et al., 2010).

[50] $EEG_w = \log\left(\frac{\sum_{i=1}^{N} \beta_i}{\sum_{i=1}^{N} (\theta_i + \alpha_i)}\right)$, While "N" is the number of channels. θ, α and β are Theta, Alpha and Beta frequency rhythms, respectively (Chanel et al., 2011).

Table 18 – Extracted EEG Features List – All Features Were Extracted From Each Window

Feature	Description	Feature	Description
14 Single Channels Theta Rhythms	4Hz to 8Hz Frequency Range Power	7 Paired Channels[51] Signal Quality	As Reported by EMOTIV EPOC, in 5 Classes
14 Single Channels Slow-Alpha Rhythms	8Hz to 10Hz Frequency Range Power	7 Asymmetric Power Ratio[49] Using Slow-Alpha Rhythms	7 Symmetric Channel Pairs
14 Single Channels Alpha Rhythms	8Hz to 13Hz Frequency Range Power	7 Asymmetric Power Ratio[49] Alpha Rhythms	7 Symmetric Channel Pairs
14 Single Channels Beta Rhythms	14Hz to 26Hz Frequency Range Power	Left Frontal EEG_w[50]	AF3, F3, F7 and FC5 Channels
14 Single Channels Gamma Rhythms	30Hz to 45Hz Frequency Range Power	Right Frontal EEG_w[50]	AF4, F4, F8 and FC6 Channels
14 Single Channels[52] Signal Quality	As Reported by EMOTIV EPOC, in 5 Classes	Left Parietal EEG_w[50]	P7 and O1 Channels
7 Paired Channels[53] Theta Rhythms	4Hz to 8Hz Frequency Range Power	Right Parietal EEG_w[50]	P8 and O2 Channels
7 Paired Channels[53] Slow-Alpha Rhythms	8Hz to 10Hz Frequency Range Power	Frontal EEG_w[50]	AF3, AF4, F3, F4, F7, F8, FC5 and FC6 Channels
7 Paired Channels[53] Alpha Rhythms	8Hz to 13Hz Frequency Range Power	Parietal EEG_w[50]	P7, P8, O1 and O2 Channels
7 Paired Channels[53] Beta Rhythms	14Hz to 26Hz Frequency Range Power	Overall EEG_w[50]	All 14 Channels
7 Paired Channels[53] Gamma Rhythms	30Hz to 45Hz Frequency Range Power	7 Signal Quality for all Measurements[54]	As Reported by EMOTIV EPOC, in 5 Classes

5.3.3.4. GSR Features

Table 19 presents all features extracted from the GSR signals. All GSR features, presented in Section 5.1.3, have been extracted from the raw GSR signal, within each window. In addition, three extra features have been introduced in the present study and have been extracted from the GSR signal: (1) mean of the positive values in the GSR first derivative, as well as negative values (employed by (Soleymani et al., 2011; Chanel et al., 2011; Koelstra et al., 2012)) have been extracted; (2) mean of the GSR first derivative's peak values (local maxima), as well as the average peaks of the original signal (employed by (Soleymani et al., 2011; Koelstra et al., 2012)) have been recorded; (3) the GSR fluctuation frequency has also been extracted, using Equation 14 ($X'(i)$ is the i^{th} element of the first derivative of the GSR signal; and

[51] Average signal quality of two symmetric channels, within the corresponding window
[52] Average channel's signal quality, within the corresponding window
[53] Voltage subtraction of two symmetric channels
[54] Average signal quality of all target channels, within the corresponding window

$sign(a)$ presents the sign function). The fluctuation frequency signifies the number of times the signal changes direction (i.e. increase to decrease and vice versa).

$$Fluctuation\ Frequency = \sum_{i=0}^{N-1} [sign(X'(i)) \neq sign(X'(i+1))]$$

Equation 14 – Signal Fluctuation Frequency Equation

The spectral power has been extracted four times, each time by using one of the power calculation formulas presented in Section 5.3.3.1 (Equation 9, Equation 10, Equation 11 and Equation 12). Consequently, in total, 14 features have been extracted from the raw GSR signals (refer to Appendix J for features indices and the corresponding values specifications – Features Indices 750 to 763).

Table 19 – Extracted GSR Features List – All Features Were Extracted From Each Window

Feature	Description	Feature	Description
Mean	Average of the GSR Signal	Mean of the Positive Values in the First Derivative	Average of the Positive Values of the First Derivative of the GSR Signal
Minimum	Minimum Value of GSR Signal	Mean of the Negative Values in the First Derivative	Average of the Negative Values of the First Derivative of the GSR Signal
Maximum	Maximum Value of GSR Signal	Mean of The First Derivative Peaks	Average of the Peak Values (Local Maximums) of the First Derivative of the GSR Signal
Standard Deviation	Standard Deviation of the GSR Signal	GSR Low Frequency Spectral Power	0Hz to 2.4Hz Frequency Range Power
Mean of The Peaks	Average of the Peak Values (Local Maximums) of the GSR Signal	Fluctuation Frequency	Frequency of the GSR Signal Direction Changing
Mean of The First Derivative	Average of the First Derivative of the GSR Signal	–	–

5.3.3.5. Heart Rate Features

Table 20 presents all features extracted from the heart rate signals. Seven out of eight heart rate features, presented in Section 5.1.3, have been extracted from the raw heart rate signal, within each window. The sampling frequency of the Shimmer+ device (512Hz) did not provide the appropriate frequency resolution required for extracting low frequency spectral power[55] from the heart rate signal. Six additional features have also been implemented in this study and were extracted from the heart rate signal: (1) minimum and (2) maximum heart rate values within a window are extracted; (3, 4) mean of both positive and negative values of the first derivative of the heart rate signals are also recorded; (5) the mean of the peak values (local maxima) of the first derivative of the heart rate is extracted from each window; (6) the heart rate fluctuation frequency is also obtained, using the algorithm presented in Equation 14 (Section 5.3.3.4). The spectral powers have been extracted four times, using one of the power calculation formulae presented in Section 5.3.3.1 (Equation 9, Equation 10, Equation 11 and Equation 12). Consequently, in total, 22 features have been extracted from the raw Heart Rate signals (refer to Appendix J for features indices and the corresponding values specifications – Features Indices 727 to 748).

Table 20 – Extracted Heart Rate Features List – All Features Were Extracted From Each Window

Feature	Description	Feature	Description
Mean	Average of the Heart Rate Signal	Mean of the Negative Values in the First Derivative	Average of the Negative Values of the First Derivative of the Heart Rate Signal
Minimum	Minimum Value of the Heart Rate Signal	Mean of The First Derivative Peaks	Average of the Peak Values (Local Maximums) of the First Derivative of the Heart Rate Signal
Maximum	Maximum Value of the Heart Rate Signal	Heart Rate Medium Frequency Spectral Power	0.04Hz to 0.15Hz Frequency Range Power
Standard Deviation	Standard Deviation of the Heart Rate Signal	Heart Rate High Frequency Spectral Power	0.15Hz to 0.5Hz Frequency Range Power
Mean of The Peaks	Average of the Peak Values (Local Maximums) of the Heart Rate Signal	Heart Rate Spectral Power Ratio[56]	Fractional Ratio of the Heart Rate Medium and High Spectral Power
Mean of The First Derivative	Average of the First Derivative of the GSR Signal	Fluctuation Frequency	Frequency of the GSR Signal Direction Changing
Mean of the Positive Values in the First Derivative	Average of the Positive Values of the First Derivative of the GSR Signal	–	–

[55] 0.01Hz to 0.04Hz as presented in (Soleymani et al., 2011; Koelstra et al., 2012; Bailenson et al., 2008; Kukolja et al., 2014; Kim & Andre, 2008; Nardelli et al., 2015).

[56] Heart Rate Spectral Power Ratio $= \frac{\text{Medium Frequency Spectral Power}}{\text{High Frequency Spectral Power}}$

5.3.3.6. Affective Tagging

Each affective feature vector F_i (each row of the features matrix, extracted from a single window) has to be tagged by the corresponding emotion (y_i), which has been experienced by the participant. As discussed in Section 5.2.2, participants were asked to self-assess their average emotional experience during each game, using both dimensional and categorical models, at the end of each sub-game. The dimensional ratings (using Valence and Arousal axes) are converted into one of the four Affective Clusters[57] (PVLA, PVHA, NVPA and NVNA – Section 2.3.5). As the self-assessments are conducted at the end of each game, rather than continuously during the gameplay, the below hypothesis has been presented in this study. According to this hypothesis, we divided the emotional experience of the participants, during a single session (game), into two affective periods:

1. **"Emotion Build-Up" Period:** This period occurs during the first part of each sub-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 sub-game or some other pre-cognitive state.
2. **"Emotion Persistence" Period:** This period occurs during the last part of each sub-game. Within this period, the emotional experience of the participant has been influenced by the current sub-game, and can be (reasonably) confidently labelled by an affective cluster or 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 and cluster, reported by the participants at the end of the sub-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. As the emotional experience of the participants can be unpredictable during the Emotion Build-Up period, 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 (refer to Appendix J for features indices and the corresponding values specifications – Features Indices 16 to 19).

5.3.3.7. Defective Data Removal

All windows exhibiting EEG signals with an average signal quality below "fair" (according to the EMOTIV EPOC signal quality classes[43]) 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.

[57] The cluster is determined according the dimensional ratings of the participants at the end of each game and the cluster boundaries, presented in Section 2.3.5.

5.4. Feature Selection

5.4.1. Minimal-Redundancy-Maximal-Relevance (mRMR)

As discussed in Section 5.3.3, a total of 743 features have been extracted from the raw physiological signals. 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 optimal features*), while they can adequately capture the essence of the data (Murphy, 2012, pp.2-12). To perform the feature selection, the minimal-redundancy-maximal-relevance (mRMR) (Peng et al., 2005) 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 optimal subset $F_S \in \mathbb{R}^{N\times d}$, such that $d \ll D$, and F_S can optimally characterise F.

The mRMR algorithm employs *Shannon's Entropy* (Murphy, 2012, p.57) 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) (Peng et al., 2005). To perform the analysis, the database has to be discretised prior to the Shannon's Entropy calculations. Therefore, all features were discretised according to 3 classes (-1, 0 and 1 – as conducted by Peng et. al.), with respect to the features' mean and standard deviation values[58] (Peng et al., 2005).

5.4.2. Feature Selection Parameters

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 can produce various lists of 30 Most optimal features, according to different windowing techniques employed in the features matrix construction process (28 different window lengths for either Hamming or Tukey windowing techniques – Section 5.3.2.4). This combination can create 1680 different settings ($2\times 28\times 30 = 1680$), for classification according to either Affective Clusters or Emotion Labels[59].

5.4.3. PCA-Correlation-Based Feature Selection

In the present study, another feature selection technique has been designed, by the present authors, and implemented to select the most optimal features. The designed feature selection technique employs Principal Component Analysis (PCA) to detect the most optimal features, which can explain majority of the dataset's variation. Moreover the technique employs a correlation-based analysis to eliminate the features, which have high dependencies, respect to each other. A detailed description of the technique has been presented in Appendix K. The mRMR technique outperforms the designed PCA-Correlation-based feature selection algorithm, by

[58] $Discretised\ F_{ij} = \begin{cases} 1 & F_{ij} > mean(F_i) + std(F_i) \\ 0 & mean(F_i) - std(F_i) \le F_{ij} \le mean(F_i) + std(F_i) \\ -1 & F_{ij} < mean(F_i) - std(F_i) \end{cases}$, while "mean" and "std" are the average and standard deviation of F_i, respectively (Peng et al., 2005).

[59] E.g. (1) most optimum feature for the 2-second Hamming window, (2) 2 most optimal features for the 2-second Hamming window, … , (1680) 30 most optimal features for the 100% Tukey window.

considerable margins (to compare the classification performances, compare Table 22 in Section 5.7.3 and Table 2 in Appendix K). Therefore, only the mRMR technique has been implemented within the present study, for the feature selection process.

5.5. Classification and Affective recognition

5.5.1. Classification Techniques

As discussed in Section 5.1.5, Support Vector Machine (SVM) (Murphy, 2012, pp.498-507), Discriminant Analysis (DA) (Murphy, 2012, pp.103-12) and Classification Trees (Murphy, 2012, pp.546-54) have been employed by the majority of the studies, reviewed. In the present study, we evaluated the performance of these classifiers (SVM, DA and Classification Tree), plus the K-Nearest Neighbour (KNN) classifier (Murphy, 2012, pp.16-18), in the affective recognition process. All classifications and cross-validations have been implemented within MATLAB software (version R2015b), using the Statistics and Machine Learning Toolbox[60].

5.5.1.1. Support Vector Machine (SVM)

Support Vector Machines are supervised classification and regression methods, originally designed for binary classifications, but with the capability for extension to be implemented in multi-class and regression applications. The SVM classifier is a Kernelized algorithm, which attempts to cluster a feature space according to a number of known labels, with maximum possible distance between the clusters' borders, by using a kernel function (Murphy, 2012, pp.498-507). There are various Kernel functions that could be implemented in SVM classification algorithms (Murphy, 2012, pp.481-88). In the present study, the Linear, 2nd Order Polynomial (Quadratic), 3rd Order Polynomial (Cubic) and Gaussian Kernel functions have been implemented and evaluated in the SVM classification. Twenty-four arbitrary Kernel Scales, for the Gaussian Kernel function, have been selected and evaluated in the cross-validation process (0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 1, 1.4, 2, 3, 4, 5 and 5.7).

5.5.1.2. Discriminant Analysis (DA)

Discriminant Analysis is a supervised discrimination (classification) algorithm, which categorises feature spaces into binary or multi-class clusters (Murphy, 2012, pp.103-12). In the present study, Linear (LDA) and Quadratic (QDA) Discriminant Analyses have been implemented and evaluated in the affective recognition process.

5.5.1.3. Classification Trees

Classification Trees (also known as Decision Trees) are supervised classification algorithms‹ which are defined by separating and partitioning a feature space, using multiple rules (called *Splits*), and defining a local model, into which feature spaces can be categorised as binary or multi-class clusters. The number of Splits in a tree can be defined as its complexity level. As an illustration, a tree with 50 Splits is five times more complex than a tree with only 10 Splits (Murphy, 2012,

[60] http://uk.mathworks.com/help/stats/index.html

pp.546-54). In the present study, twenty different arbitrary Splits numbers have been selected and evaluated in the cross-validation process (5 to 100 with step size of 5).

5.5.1.4. K-Nearest Neighbour (KNN)

K-Nearest Neighbour is a supervised classification algorithm, which categorises feature spaces into binary or multi-class clusters. To do so, the algorithm employs a training dataset to classify further data points according to the K closest data points in the training dataset (K nearest neighbours – using Euclidean distance[61]) (Murphy, 2012, pp.16-18). In the present study, 30 different arbitrary K values have been implemented and evaluated in the cross-validation process (1 to 30).

5.6. Hyper-Parameters Tuning

As explained in Section 5.3.2.4, 56 different windowing settings can be applied (each of which with different length and type) to generate 56 various feature matrices. The feature selection algorithm, presented in Section 5.4, is capable of identifying different sets of the most optimal features, by employing different feature matrices constructed from various window lengths and types, to classify the physiological affective space. On the other hand, the feature selection algorithm, presented in Section 5.4, can generate 30 feature sets, each of which containing between 1 and 30 of the most optimal features. This will result in 1680 different training matrices for the classification process (Section 5.4.2). By employing the Affective Clusters (4 clusters – dimensional assessment) and Emotion Labels (8 labels – categorical assessment), the classifiers could categorise the physiological responses into four or eight classes, respectively (Section 5.2.1.1).

The performance of the classifiers, trained according to each training matrix, vary in terms of the classification accuracy. These variables are the *hyper-parameters* of the affective recognition system. Hyper-parameters are the restrictions of learning algorithms, which can be tuned prior to the training process, resulting in various classification performances (Bergstra & Bengio, 2012). The process, in which the best set of hyper-parameters, which can produce the best classification performance, is identified, is called *hyper-parameters tuning* (Bergstra & Bengio, 2012). In machine learning algorithms, cross-validation is mainly used as a measure in the hyper-parameters tuning process. However, various searching algorithms (e.g. grid or random search) are employed in addition, in order to identify the best set of hyper-parameters, which generate the best classification accuracy (Bergstra & Bengio, 2012).

It is almost impossible to train each classifier according to all possible hyper-parameter variations, as there are infinite combinations (considering all possible window lengths, rather than a finite number of arbitrary selections). Therefore, by selecting the arbitrary window lengths (Section 5.3.2.4), the number of required features (Section 5.4.2) and the classification settings (Section 5.5.1.1, Section 5.5.1.2, Section 5.5.1.3 and Section 5.5.1.4), the number of possible hyper-parameters

---

[61] $Euclidean\ Distance = \sqrt{\sum_{i=1}^{D}(f_{1i} - f_{2i})^2}$, while f_{1i} and f_{2i} are the i^{th} features of the 1st and 2nd points respectively. And "D" is the number of features, employed to present each point in the affective space.

tuning variations have reached to 132,720[62] for each classification process (according to either Affective Clusters or Emotion Labels – 265,440 in total). According to (Bergstra & Bengio, 2012), this small subset of the infinitely larger settings space may be sufficient in the hyper-parameters tuning process, as the majority of the hyper-parameters variations do no matter much, as only those which result in high accuracy actually matter. To assess the performance of different classifications settings, the accuracies of the classifiers have been estimated through a 10-Fold (random folding) Cross-Validation technique (Murphy, 2012, p.209). To train and cross-validate the performance of all 265,440 classifiers, a processing farm service (HPC[63]) was employed to speed up the process.

To perform the comparison, the scatter plots of the classification accuracies, for each setting, have been analysed. As an illustration, the scattered dots in Figure 22 define different classifiers with various settings. For example, if the classifier employs five features for the classification, different window types (Hamming *vs*. Tukey), window lengths (28 different window lengths), classifier settings (different K-value in KNN, etc.), and so on, 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 settings, which generate the maximum classification accuracy is identified and highlighted (e.g. the line, highlighting the maximum values in Figure 22). The analyses of different hyper-parameters variations are presented in Sections 5.6.1, 5.6.2 and 5.6.3. The best performing classifiers (with the highest accuracy) have been identified and presented in Section 5.7.1.

5.6.1. Number of Features Evaluation

Figure 22 and Figure 23 present the performance of the classifiers with respect to the different number of features, according to Affective Clusters and Emotion Labels, respectively. As it can be obtained by the graphs, 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 DA performance has not been changed considerably, by employing more or less features, whereas employing more features has increased the accuracy of the Classification Tree. The accuracy of both KNN and SVM classifiers, with respect to the number of employed features, increases by employing more features, while saturating around 98%. As it can be seen in the graphs, the accuracy of KNN and SVM 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.

[62] $(1680{\times}27)_{SVM} + (1680{\times}2)_{DA} + (1680{\times}20)_{Classification\ Tree} + (1680{\times}30)_{KNN} = 132{,}720$

[63] High Performance Computing (HPC) services provided by Queen Marry University of London (QMUL): http://docs.hpc.qmul.ac.uk/

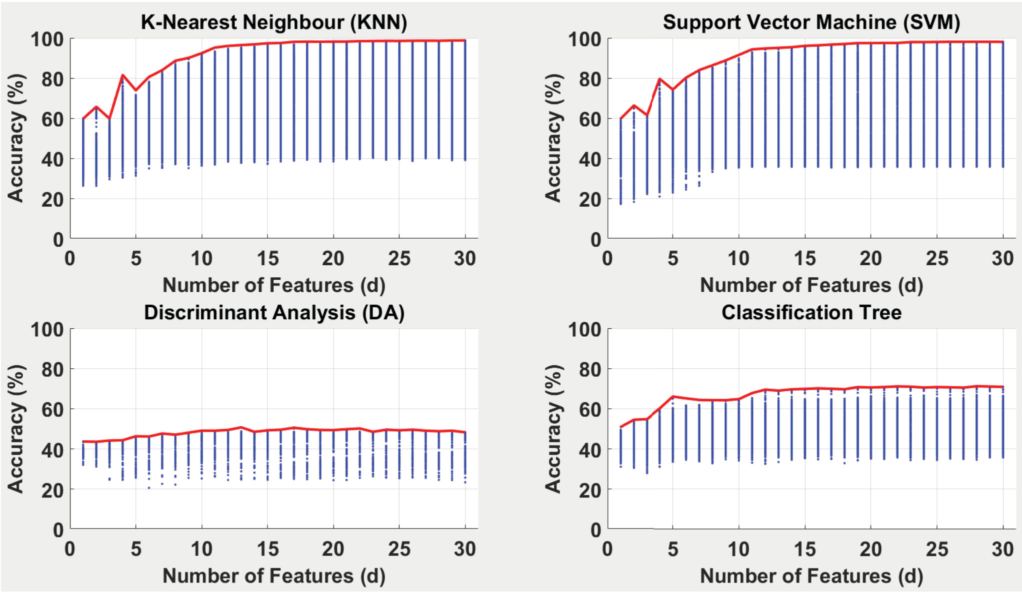


Figure 22 – Classifiers Performance (Cross-Validation) Respect to Different Number of Features, According to Affective Clusters

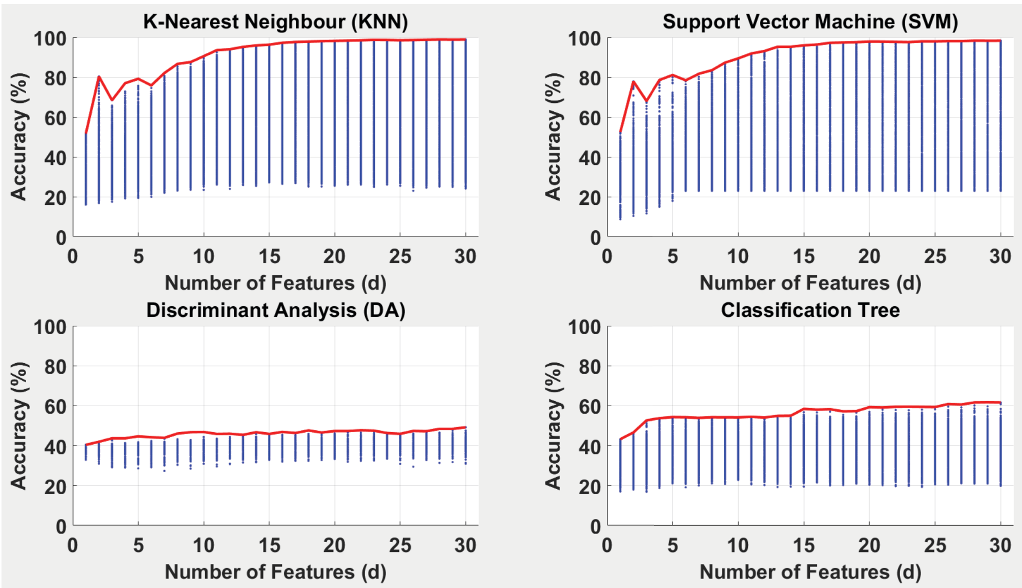


Figure 23 – Classifiers Performance (Cross-Validation) Respect to Different Number of Features, According to Emotion Labels

5.6.2. Windowing Settings Evaluation

As discussed in Section 5.3.2.4, there are two tuning parameters for the windowing process; window type (Hamming *vs.* Tukey) and length (17 fixed *vs.* 11 relative). As shown in Figure 24 and Figure 25, the performances of KNN, SVM and Classification Tree (according to both Affective Clusters and Emotion Labels) are slightly better when using the Hamming window, compared to the Tukey window. The DA classifier performance did not change considerably by employing either the Hamming or Tukey window.

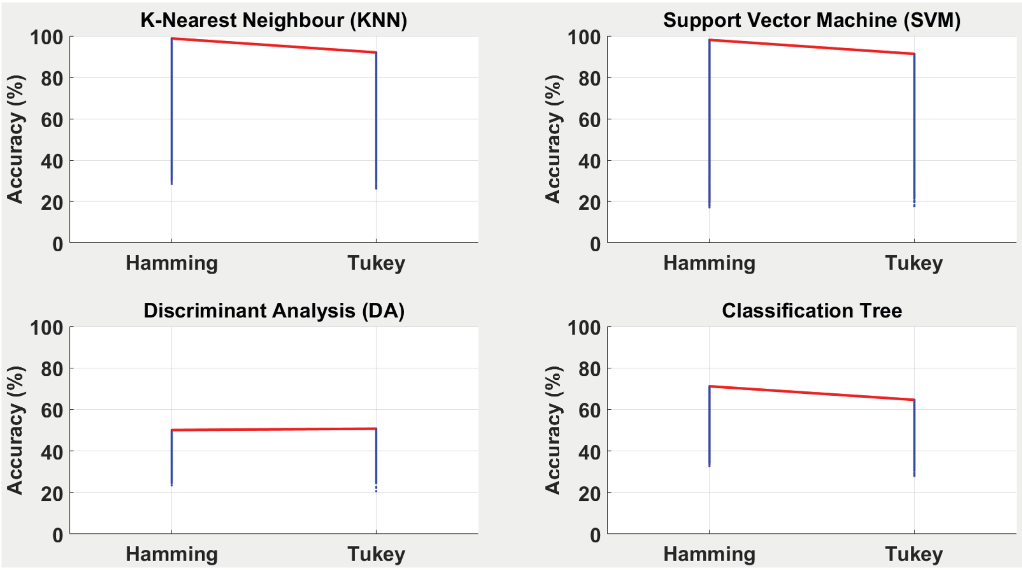


Figure 24 – Window Type Vs. Classifiers Accuracy (Cross-Validation), According to Affective Clusters

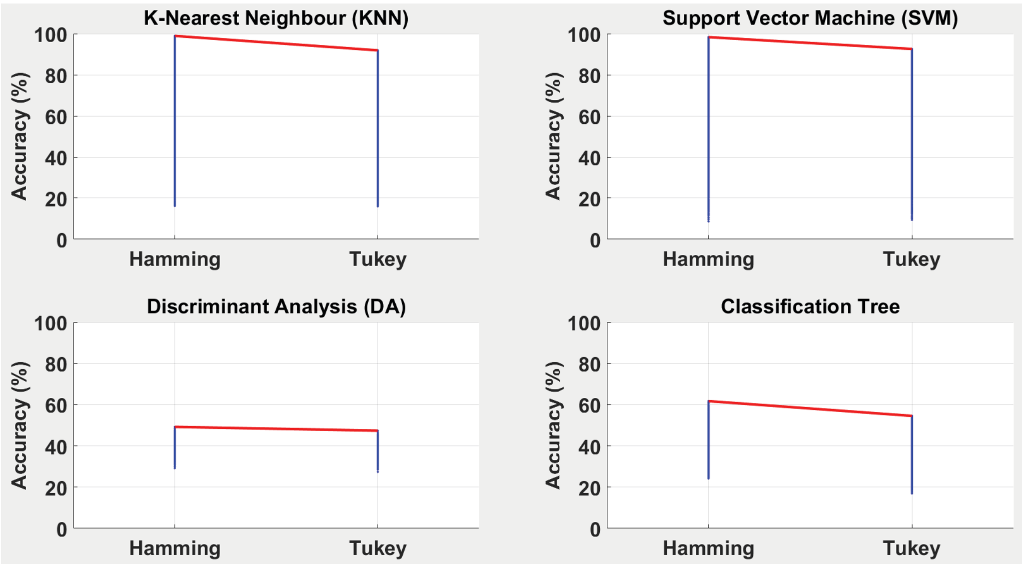


Figure 25 – Window Type *vs.* Classifiers Accuracy (Cross-Validation), According to Emotion Labels

Figure 26 and Figure 27 present the classification accuracies of the classifiers, respect to different **fixed** window lengths, according to Affective Clusters and Emotion Labels, respectively. Figure 28 and Figure 29 on the other hand, present the classification accuracies of the classifiers, with respect to different **relative** window lengths, according to Affective Clusters and Emotion Labels, respectively. As it can be seen in the graphs the performance of KNN, SVM and Classification Tree classifiers have been increased by employing shorter windows (in both fixed and relative windowing techniques). However, the DA performance has not been changed considerably by using different window lengths (except for the relative windowing technique, in classification respect to Emotion Labels).

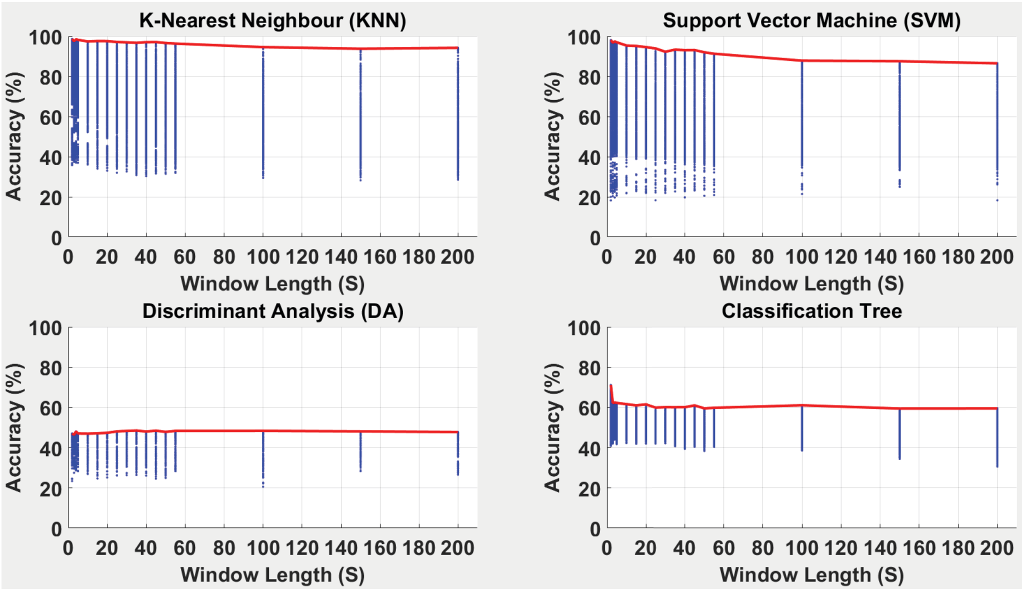


Figure 26 – Window Fixed Length Vs. Classifiers Accuracy (Cross-Validation), According to Affective Clusters

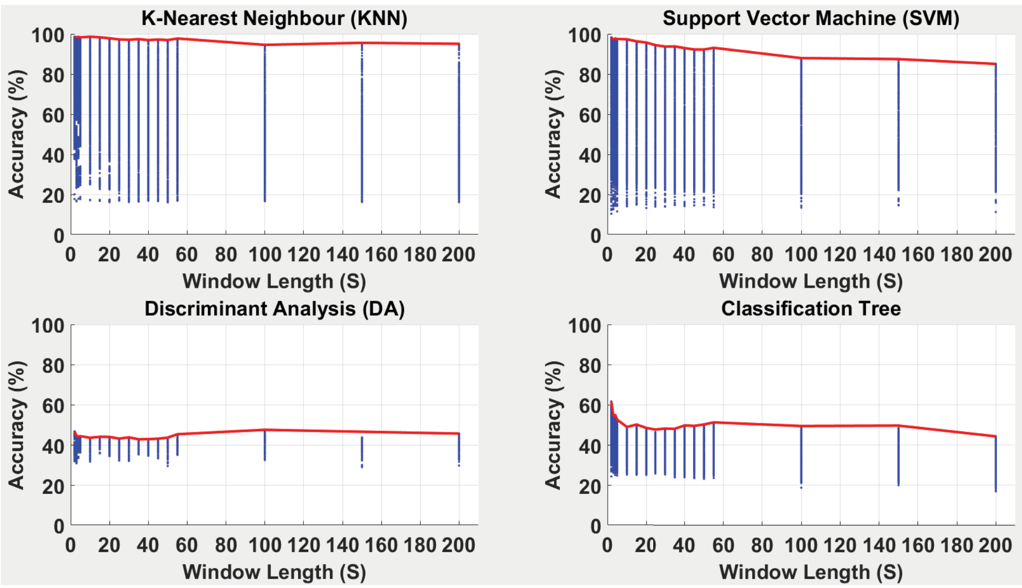


Figure 27 – Window Fixed Length Vs. Classifiers Accuracy (Cross-Validation), According to Emotion Labels

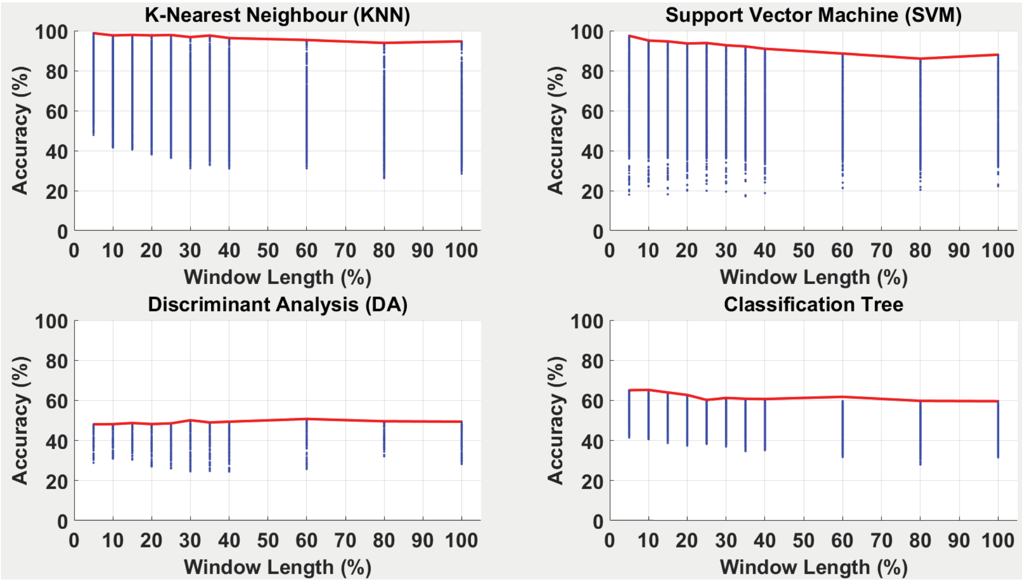


Figure 28 – Window Relative Length Vs. Classifiers Accuracy (Cross-Validation), According to Affective Clusters

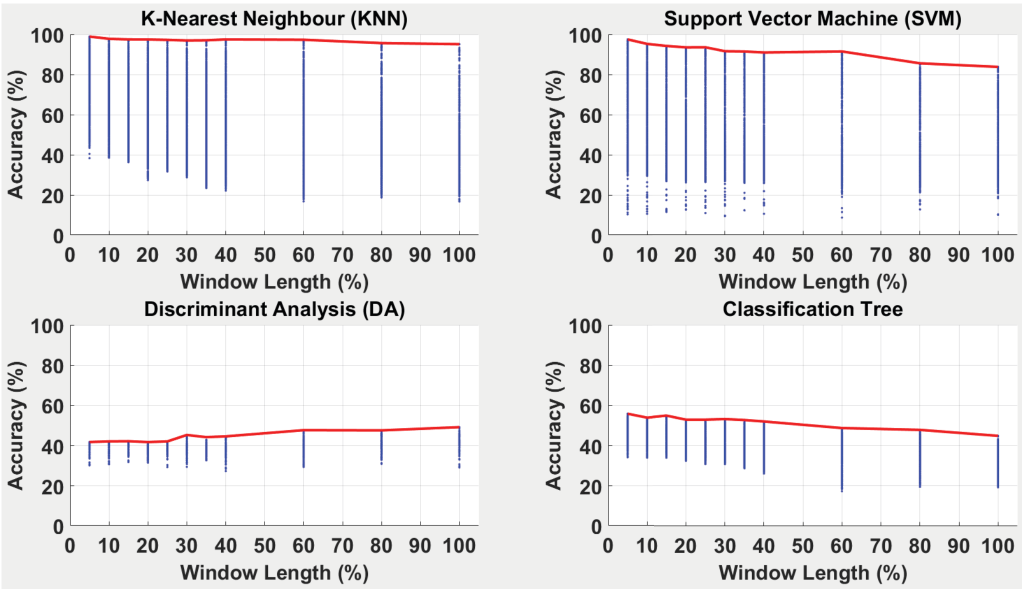


Figure 29 – Window Relative Length Vs. Classifiers Accuracy (Cross-Validation), According to Emotion Labels

5.6.3. Classifiers Settings Evaluation

As discussed in Section 5.5.1, each classifier can be tuned according to a parameter. As can be seen in Figure 30 and Figure 31, the performance of the KNN classifier is slightly attenuated, whilst "K" is increased. 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. In contrast, as shown in figures, the accuracy of the Classification Tree is boosted, while the number of Splits is increased. This means that the Classification Tree performs better, if more conditions are defined and a more complex tree is generated. According to the analysis, the Classification Tree has its maximum accuracy if 100 Splits are generated in the Tree. On the other hand, the performance of the DA classifier is attenuated in Affective Clusters if a 2nd order polynomial (quadratic instead of linear) function is employed; whereas it did not change considerably, by employing linear or quadratic functions, in Emotion Labels.

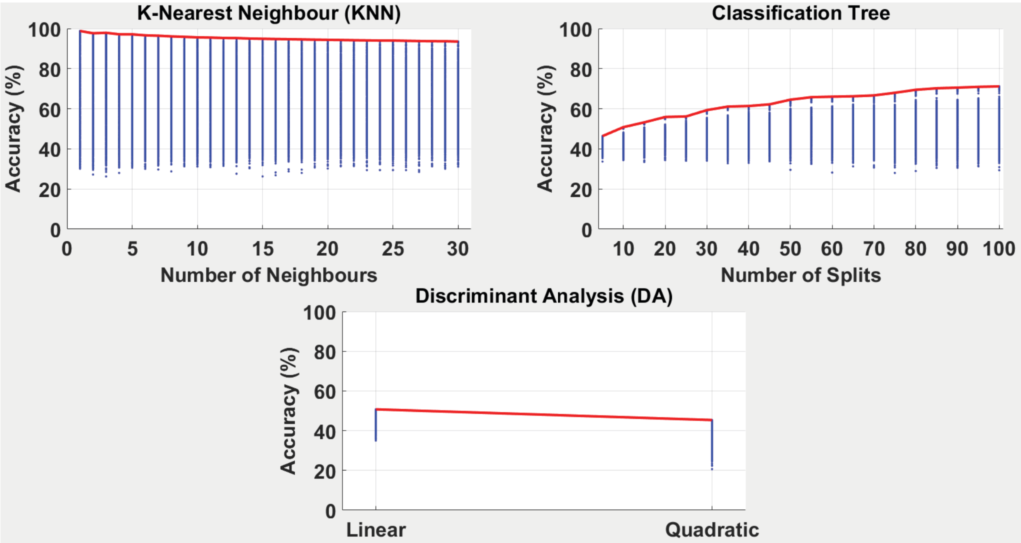


Figure 30 – KNN, Classification Tree and DA Classifiers Settings *vs.* Accuracy (Cross-Validation), According to Affective Clusters

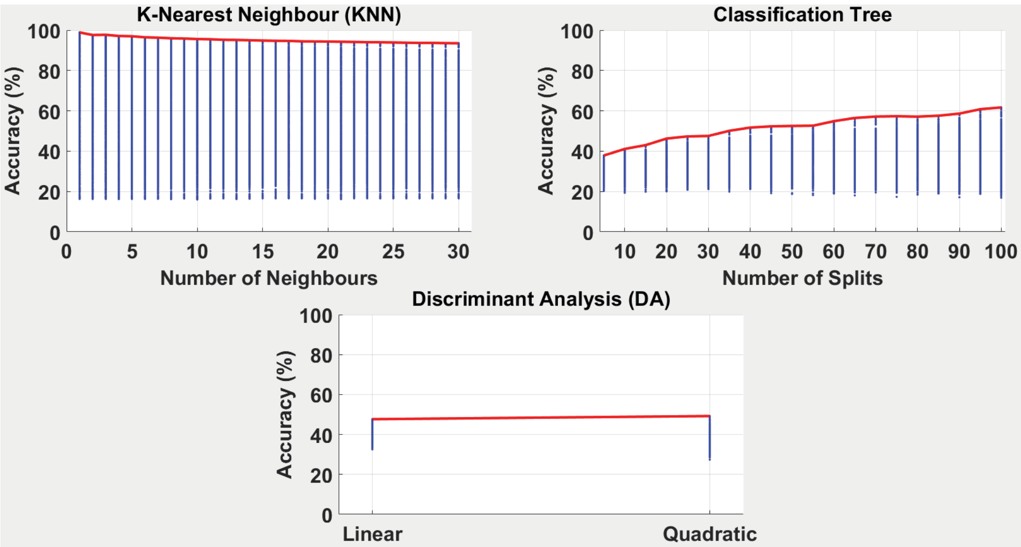


Figure 31 – KNN, Classification Tree and DA Classifiers Settings Vs. Accuracy (Cross-Validation), According to Emotion Labels

Figure 32 and Figure 33 present the performance of the SVM classifier, according to four different Kernel functions; linear, 2nd order polynomial (quadratic), 3rd order polynomial (cubic) and Gaussian function with 24 different Kernel scales. As illustrated by the graphs, 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. The Cubic Kernel performance was very similar to the Gaussian function, but only 10% of the best performing classifiers (Section 5.7.1) employed the Cubic Kernel.

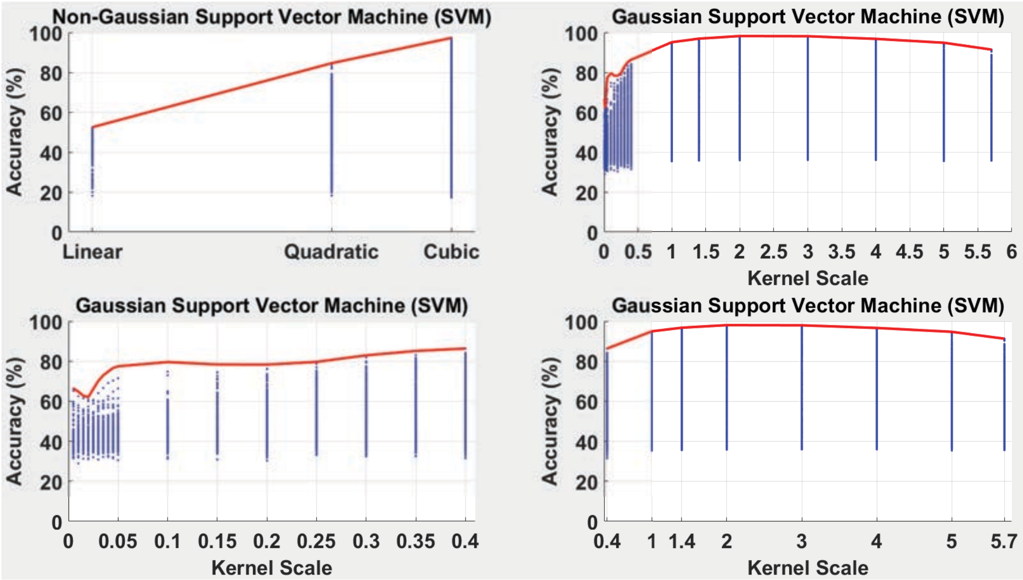


Figure 32 – SVM Classifier Settings Vs. Accuracy (Cross-Validation), According to Affective Clusters – The Right Top Graph Presents the *Kernel* Scale Between 0.005 and 5.7 – The Bottom Graphs Presents Magnified Versions, with *Kernel* Scale Between 0.005 and 0.4 (Left) and 0.4 and 5.7 (Right)

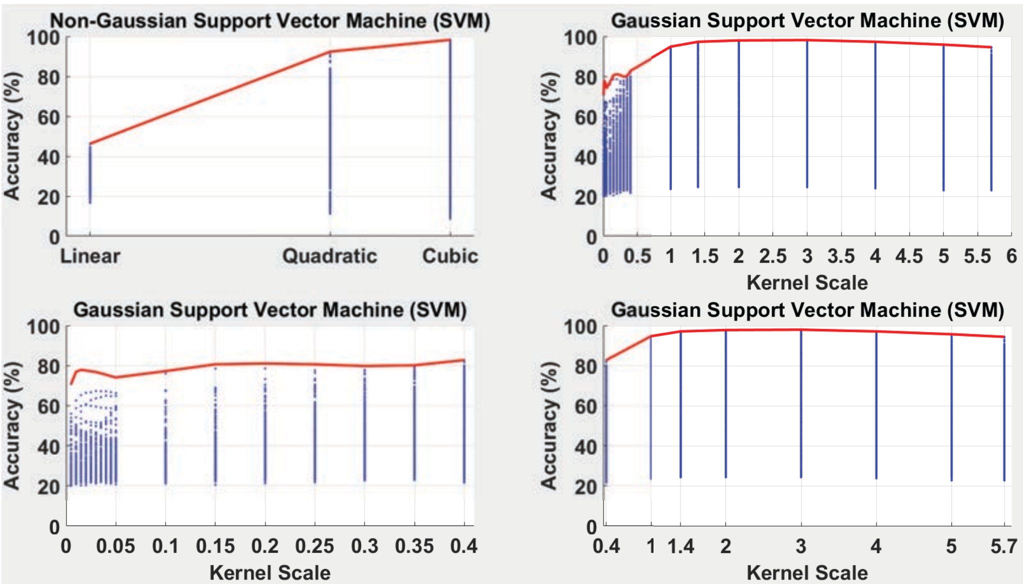


Figure 33 – SVM Classifier Settings Vs. Accuracy (Cross-Validation), According to Emotion Labels – The Right Top Graph Presents the *Kernel* Scale Between 0.005 and 5.7 – The Bottom Graphs Presents Magnified Versions, with *Kernel* Scale Between 0.005 and 0.4 (Left) and 0.4 and 5.7 (Right)

5.7. Discussion

5.7.1. Best Performing Classifiers

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, SVM, DA and classification tree) have been identified. Figure 34 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)[64] showed that the different windowing techniques (fixed *vs.* relative) is not a significant factor in changing the classification performances ($P_{Windowing} = 0.691$). However, the performances of different classification techniques are significantly different, in terms of their classification accuracy ($P_{Classification} < 0.001$). The KNN (96.78% (±1.42%) and 97.24% (±1.14%) mean accuracy, for Affective Clusters and Emotion Labels respectively, across all window lengths) and SVM (92.77% (±3.42%) and 92.91% (±3.97%) mean accuracy, for Affective Clusters and Emotion Labels respectively, across all window lengths) perform better than the Classification Tree (61.54% (±2.46%) and 51.06% (±3.62%) mean accuracy, for Affective Clusters and Emotion Labels respectively, across all window lengths). The DA classifier performs worst than the other three, with 48.28% (±0.96%) and 44.53% (±1.98%) mean accuracy, for Affective Clusters and Emotion Labels respectively, across all window lengths. But all performed significantly better than random classifiers.

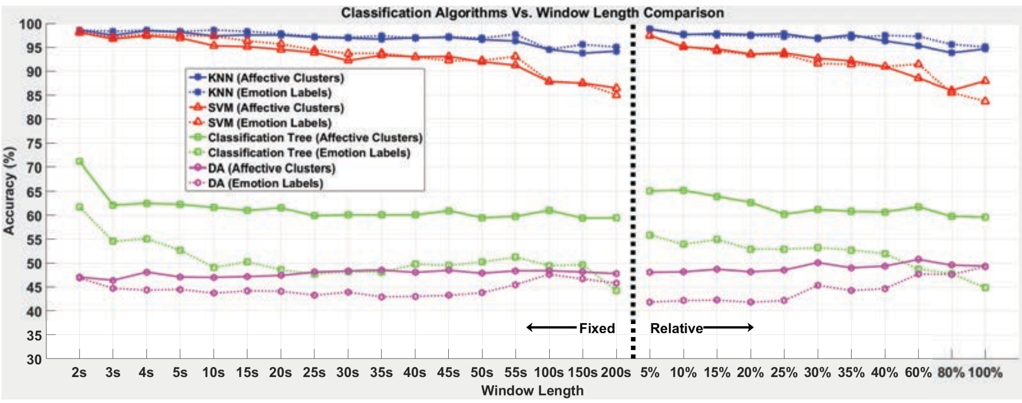


Figure 34 – Classification Methods Comparison (Cross-Validation), According to Affective Clusters and Emotion Labels – The Horizontal Axes Presents 28 Different Window Lengths; 17 Fixed (Left Side of the Vertical Dashed Line) and 11 Relative (Right Side of the Vertical Dashed Line)

As there is a high and significant negative correlation between the window durations and the accuracies of the KNN and SVM classifiers ($r_{Fixed}(68) = -0.69$ and $r_{Relative}(44) = -0.61$, P<0.001 – for both Affective Clusters and Emotion Labels), one could conclude that, by decreasing the window length, the accuracy of the KNN and SVM classifiers can be improved. In the fixed windowing technique,

[64] Classifiers accuracy is considered as the dependent variables, while relative vs. fixed windowing technique and different classifiers as the independent parameters.

the KNN classifier achieved its best performance by the 2-second window with 98.61% accuracy, for both Affective Clusters and Emotion Labels. The SVM classifier achieved its best performance using the same 2-second window, with 98.07% and 98.30% accuracy, for Affective Clusters and Emotion Labels, respectively. In the relative windowing technique on the other hand, the KNN and SVM classifiers achieved their best performance by the 5% window with 98.7% and 97.47% accuracy, for Affective Clusters and Emotion Labels, respectively.

5.7.2. Most Optimal features

As discussed in Section 5.4, 743 features were extracted from the recorded physiological signals, while no more than the 30 most optimal features were selected, for the classification process, using the mRMR algorithm. There are 56 different windowing techniques (different length and type – Section 5.3.2.4), which result in 56 various sets of most optimal features. Furthermore, depending on the classification class (either Affective Clusters of Emotion Labels) provided for the mRMR algorithm, different sets of most optimal features could be nominated (Section 5.4). The mRMR algorithm guarantees to find the *d* most optimal features, which have minimum mutual information amongst each other (minimum redundancy), whilst maximum mutual information with respect to the classification classes (maximum relevance – Affective Clusters or Emotion Labels). Therefore, in total, 112 different sets of the most optimal features[65] (each of which containing 30 features – Section 5.4.2) have been identified by the mRMR algorithm.

On the other hand, to consider the performance of classifiers as well (according to the cross-validation accuracies), the sets of most optimal features, extracted from the Tukey-based windows, were excluded[66]. Therefore, 56 different sets of the most optimal features were analysed and the *unique features*, which are present in at least one of the 56 sets, were identified. This *Unique Most Optimal features List* contains 250 features, out of the 743 features, identified at the beginning of the process. Furthermore, by considering the best performing KNN and SVM classifiers[67] in each window, the features, which have not been employed in the classification process, have been excluded, and the number of features in the Unique Most optimal features List has been reduced to 230 features. This guarantees that these sets of optimum features are those, which not only preserve the maximum relevance and minimum redundancy aspect of the mRMR results, but also generate the best classification accuracies, according to both Affective Clusters and Emotion Labels.

Analysing this list can highlight the superiority of each feature against the others. To do so, the popularity of all features of the Unique Most optimal features List (containing the 230 features), within the 56 different sets of most optimal features[68], have been calculated. This value has been reported as a percentage, signifying the portion of windows (among all 56 windows) that employed a particular

[65] 28 sets for Hamming and 28 sets for Tukey windows, for both Affective Clusters and Emotion Labels ($28\times4 = 112$).
[66] All classifiers with the highest accuracy (Section 5.7.1) employed Hamming window technique.
[67] The KNN and SVM classifiers outperformed the Classification Tree and DA classifiers, considerably (Section 5.7.1), so only these 2 were considered in this process.
[68] 28 Hamming windows for classification according to Affective Clusters and 28 according to the Emotion Labels.

feature. This analysis has **not** been used in the classification process, and has only been developed for appropriate presentation purposes. Table 21 presents the Unique Most optimal features List, grouped according to their measurement categories. The table presents the frequency percentages as well, and these signify the occurrence frequency ranges in which each feature group has been employed within a windowing and classification technique (among all 56 windows).

Table 21 – Selected Features According to the mRMR algorithm and the Best Performing Classifiers

<table>
<tr><th>Feature Group</th><th colspan="3">Detail</th><th colspan="3">Frequency Percentage
(25^{th} – 75^{th} Percentiles)</th></tr>
<tr><td>Age</td><td colspan="3">Participants age according to 4 classes</td><td colspan="3">100%</td></tr>
<tr><td>GSR</td><td>Minimum</td><td>Mean</td><td>Fluctuation Frequency</td><td>100%</td><td>89.29%</td><td>67.86%</td></tr>
<tr><td>Heart Rate</td><td colspan="3">Maximum</td><td colspan="3">55.36%</td></tr>
<tr><td>Alpha Rhythm Asymmetric Ratio</td><td colspan="3">All 7 paired channels</td><td colspan="3">43.82% (21.42% - 65.62%)</td></tr>
<tr><td>Heart Rate</td><td colspan="3">Fluctuation Frequency</td><td colspan="3">39.29%</td></tr>
<tr><td>Slow-Alpha Rhythm Asymmetric Ratio</td><td colspan="3">All 7 paired channels</td><td colspan="3">38.52% (25.44% - 54.91%)</td></tr>
<tr><td>EEG Gamma Rhythms</td><td colspan="3">6 paired channels (Excluding F7-F8)
8 single channels (Excluding AF3, F3, F4, F7, F8, P8)</td><td colspan="3">33.93% (7.14% - 58.92%)</td></tr>
<tr><td>Heart Rate</td><td colspan="3">Mean of the Peaks</td><td colspan="3">30.36%</td></tr>
<tr><td>Hand Preference</td><td colspan="3">Participants dominant hand</td><td colspan="3">28.57%</td></tr>
<tr><td>EEG$_w$</td><td colspan="3">4 paired channels (Excluding F7-F8, FC5-FC6, O1-O2)
7 single channels (Excluding F3, F4, F7, F8, FC6, T8, P8)</td><td colspan="3">28.57% (8.48% - 43.75%)</td></tr>
<tr><td>GSR</td><td colspan="3">Low Frequency Power</td><td colspan="3">23.21%</td></tr>
<tr><td>Heart Rate</td><td colspan="3">Medium Frequency Power</td><td colspan="3">23.21%</td></tr>
<tr><td>EEG Theta Rhythm</td><td colspan="3">5 paired channels (Excluding F3-F4, T7-T8)
3 single channels (AF4, P7, O2)</td><td colspan="3">22.77% (6.25% - 26.78%)</td></tr>
<tr><td>EEG Alpha Rhythm</td><td colspan="3">5 paired channels (Excluding F3-F4, P7-P8)
8 single channels (Excluding AF3, AF4, F3, F4, T7, P8)</td><td colspan="3">20.19% (8.48% - 26.78%)</td></tr>
<tr><td>Alpha-Beta Ratio</td><td colspan="3">All 7 paired channels
8 single channels (Excluding F3, F4, F8, FC6, T7, O1)</td><td colspan="3">17.38% (9.37% - 19.64%)</td></tr>
<tr><td>EEG Beta Rhythm</td><td colspan="3">6 paired channels (Excluding T7-T8)
7 single channels (Excluding AF3, F3, F7, F8, FC5, T8, P8)</td><td colspan="3">15.93% (3.57% - 30.35%)</td></tr>
<tr><td>Heart Rate</td><td colspan="3">Power Spectral Ratio</td><td colspan="3">12.5%</td></tr>
<tr><td>Gender</td><td colspan="3">Male or Female</td><td colspan="3">10.71%</td></tr>
<tr><td>Heart Rate</td><td colspan="3">Mean</td><td colspan="3">10.71%</td></tr>
<tr><td>EEG Slow-Alpha Rhythm</td><td colspan="3">6 paired channels (Excluding FC5-FC6)
6 single channels (AF3, AF4, T8, P7, O1, O2)</td><td colspan="3">10.71% (2.67% - 8.92%)</td></tr>
<tr><td>GSR</td><td colspan="3">Mean of the first derivative</td><td colspan="3">3.57%</td></tr>
<tr><td>Heart Rate</td><td>Minimum</td><td colspan="2">High Frequency Power</td><td colspan="2">3.57%</td><td>1.79%</td></tr>
</table>

Among participant-related features, “Age” has been employed by all windows, as the most optimum feature, to classify the participants’ emotional experiences (according to both Affective Clusters and Emotion Labels). This signifies the fact that the participants’ age can provide a substantial amount of information, to classify their emotional experiences. This relationship has been addressed by other studies, as well (Section 4.3.1.4).

On the other hand, and as discussed in Section 5.3.3.1, four different calculation techniques have been employed in performing the spectral analysis. In Table 21 the rhythms measured with various techniques have been combined. As an illustration, the alpha rhythm asymmetric ratio for the AF3-AF4 paired electrodes has been employed by 68% of the windowing and classification techniques, whilst 17.5% measured the powers through Summation (Equation 9), 3% based on the Power Ratio (Equation 10), 1.5% according to the RMS Ratio db (Equation 12) and the remaining 46% through the RMS (Equation 11) technique. To compare the frequency percentages of the spectral measurements according to all 4 techniques (Equation 9, Equation 10, Equation 11 and Equation 12), an Analysis of Variance (ANOVA)[69] was conducted. The analysis showed that there is no statistical difference between the spectral power calculation techniques ($P_{Measurment\_Technique} = 0.542$). One can conclude that, no spectral analysis technique is superior when compared to the others.

5.7.3. Affective Clusters vs. Emotion Labels Classification

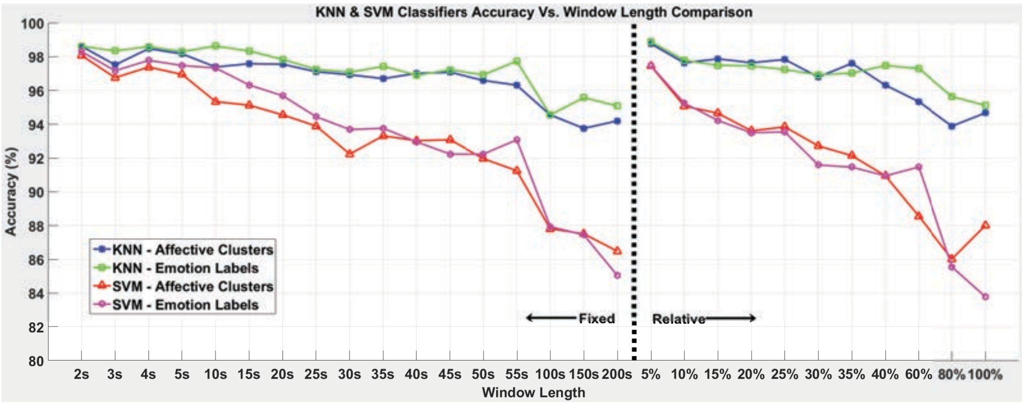


Figure 35 – Accuracy (Cross-Validation) Comparison of KNN and SVM Classifiers, According to Both Affective Clusters and Emotion Labels – Refer to Appendix L for the Classifiers Specifications

Figure 35 presents the accuracies of KNN and SVM classifiers, while classifying the features space with respect to Affective Clusters and Emotion Labels. An Analysis of Variance (ANOVA)[70] showed that the performance of the classifiers, in categorising the emotions into either Affective Clusters or Emotion Labels is not statistically different ($P_{Clusters}$=0.569). However, the performances of KNN and SVM classifiers are significantly different, in terms of their classification accuracy ($P_{Classification}$<0.001). On average, KNN (97.01% (±1.3%) mean accuracy across different windowing techniques, Affective Clusters and Emotion Labels)

[69] The popularity percentage as the dependent variable, while the measurement technique as the independent parameter.
[70] Classifiers accuracy is considered as the dependent variables, while different classifiers and Affective Clusters vs. Emotion Labels classification technique as the independent parameters.

outperformed the SVM algorithm (92.84% (±3.67%) mean accuracy across different windowing, Affective Clusters and Emotion Labels) with around 4% (Refer to Appendix L for the classifiers specifications).

As well as the classification accuracy, the F1-Score is another measure that could evaluate the performance of a classifier. Equation 15 presents the F1-Score formula, which is the harmonic mean of *Precision* (the fraction of the identified instances that are relevant), and *Recall* (the fraction of relevant instances that are identified) (Murphy, 2012, pp.184-86). The accuracy alone cannot accurately evaluate the performance of classifiers. As an illustration, consider a database containing 990 samples of class "A" and 10 of class "B". Now consider a classifier, which identifies all samples as class "A" in that database. The accuracy of this classifier (which does not do anything at all) is 99%. However, its F1-Scores for class "A" and "B" are 99.5% and 0%, respectively. The average F1-Score for both classes is 49.75%, which is a better measurement to evaluate the classifier's performance, when compared to the accuracy measurement. Therefore, the F1-Score is also employed to assess the classifiers' performance in more detail.

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Equation 15 – F1-Score Equation (Murphy, 2012, pp.184-86)

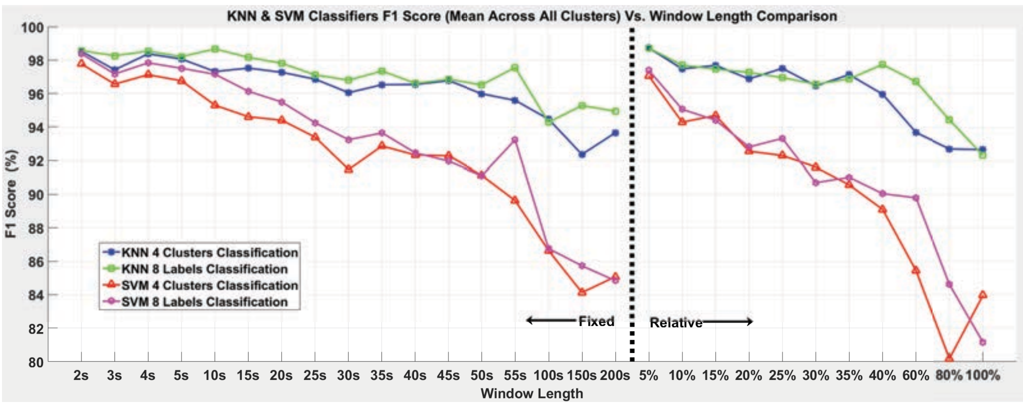


Figure 36 – KNN and SVM Mean F1-Score (Cross-Validation) Across All Clusters – Refer to Appendix L for the Classifiers Specifications

Figure 36 presents the KNN and SVM classification F1-Score, averaged across classes[71]. No game in the experiment has been able to evoke sadness on the part of the participants. Therefore, the classifiers, trained according to Emotion Labels, have not been able to classify any part of the features space into the "Sad" Label. To be able to compare the performance of the classifiers (with respect to their F1-Score), an Analysis of Variance (ANOVA)[72] has been conducted. The analysis highlighted a significant difference in F1-Score generated by different classifiers ($P_{Clasifier} < 0.001$) and classification according to either Affective Clusters or Emotion Labels ($P_{Classification\text{-}Class} < 0.001$). This means that the classifiers' F1-Score, in categorising the

[71] F1-Score is calculated within each class. Therefore in each windowing technique, 4 and 8 F1-Score for each (respectively) Affective Cluster, and Emotion Label, are calculated.
[72] The F1-Score as the dependent variable, while the classification class and the classifier as the independent parameters.

emotions into either Affective Clusters or Emotion Labels, is statistically different (unlike the accuracy analysis). Table 22 presents the mean F1-Scores for Affective Clusters, Emotion Labels and classifiers. On average, KNN outperformed the SVM classifier, while the classification according to Affective Clusters performed better, when compared to Emotion Labels (Refer to Appendix L for the classifiers specifications).

Table 22 – Mean F1-Scores (Cross-Validation) Across Classifiers, Classes and All Windowing Techniques – (A - B) Presents the (A) 25th and (B) 75th Percentile – Refer to Appendix L for the Classifiers Specifications

Emotion Label	**Mean F1-Score (25th – 75th Percentiles)**	**Affective Cluster**	**Mean F1-Score (25th – 75th Percentiles)**
Relaxed	95.50% (94.31% - 98.07%)	PVLA	94.73% (92.81% - 97.56%)
Content	95.06% (93.78% - 97.74%)		
Happy	94.12% (92.50% - 97.27%)	PVHPA	95.35% (94.34% - 97.34%)
Excited	95.61% (94.04% - 97.94%)		
Angry	94.40% (91.69% - 97.63%)	NVPA	94.82% (93.28% - 97.42%)
Afraid	93.11% (89.86% - 97.25%)		
Sad	Not Available	NVNA	90.76% (87.78% - 96.12%)
Bored	94.87% (93.64% - 97.59%)		

Emotion Labels Classifier	**Mean F1-Score (25th – 75th Percentiles)**	**Affective Clusters Classifier**	**Mean F1-Score (25th – 75th Percentiles)**
KNN	96.94% (96.23% - 98.09%)	KNN	96.29% (95.29% - 97.66%)
SVM	92.40% (89.81% - 96.19%)	SVM	91.54% (88.70% - 95.42%)

5.8. Conclusion

In this chapter the steps of designing and evaluation (cross-validation) of an affective computing system have been discussed in detail (research question III presented in Section 1.8). By conducting a deep literature review, various signal recording, pre-processing, windowing, feature extraction and classification techniques, employed by 30 affective computing studies, conducted since 1993, have been investigated (Section 5.1). Then by conducting an experiment, the psychophysiological responses of 30 participants, exposed to the Affective VR, have been recorded (Section 5.2). Various features, pre-processing, windowing and classification techniques under different tuning settings have been employed to train more than a quarter of a million different classifiers (Sections 5.3, 5.4 and 5.5). By conducting Hyper-Parameter tuning, 112 best performing classifiers, under 28 different windowing techniques, have been selected (Figure 35, Figure 36 and Table 22 – Section 5.6 and 5.7). The results suggested that the KNN and SVM classifiers outperformed the DA and Classification Tree classifiers. The KNN classifiers achieved 96.78% (±1.42% – 98.5% maximum accuracy) and 97.24% (±1.14% – 98.5% maximum accuracy) mean accuracies, for Affective Clusters and Emotion

Labels respectively (across all window lengths), while the SVM classifiers reached 92.77% (±3.42% – 98% maximum accuracy) and 92.91% (±3.97% – 98% maximum accuracy) mean accuracies, for Affective Clusters and Emotion Labels respectively (across all window lengths).

As discussed in Section 1.7, three (out of six) previous studies, which employed virtual environments as their affective stimulation medium, have conducted affective classification (Liu et al., 2009; Wu et al., 2010; Chanel et al., 2011); while the others only investigated the brain's regional activity variation patterns, under different emotional experiences (Reuderink et al., 2013; Parnandi et al., 2013; Rodríguez et al., 2015). Liu et al managed to achieve 88.9% cross-validation accuracy, according to 3 clusters (low, medium and high anxiety) (Liu et al., 2009), while Wu et al achieved 84% accuracy in a subject-dependent setting and random classifications in a subject-independent setting, according to 3 clusters (low, medium and high arousal level) (Wu et al., 2010). Chanel et al achieved 63% classification accuracy, according to 4 clusters (combining low and high pleasure and arousal levels) (Chanel et al., 2011).

The purpose of this chapter was to design an affective computing system, capable of detecting and classifying human emotions through psychophysiological behaviours, with high precision. In this chapter we were able to design and train multiple affective computing systems with maximum classification accuracy of 98.5%. This maximum cross-validation accuracy is 2% higher than the maximum cross-validation accuracy, achieved by the affective computing studies conducted since 1993 (refer to Section 5.1.5). Moreover the maximum accuracy of the designed classifiers (according to cross-validation) is much higher than the cross-validation accuracy achieved by the previous studies that conducted affective classification within virtual environments.

The most important contribution of this chapter is the employment of all affective computing techniques, which are implemented within different steps of the classification process, by various studies in the past 25 years (e.g. different windowing and classification techniques, all features employed in various studies, etc.), and integrating them within more than a quarter of a million different classifiers, to identify the best performing affective classifiers. We believe that this contribution not only summarises the breadth of research over the past 25 years, but also serves to clarify various significant aspects and details of this increasingly valuable and relevant research area.

Another contribution of this chapter is the implementation of affective computing systems within virtual environments. As mentioned earlier, only a minority of the affective computing studies conducted since 1993 have employed VR in the emotion stimulation processes (considering the fact that half of those studies employed 2D Retro games, rather than 3D environments). The resurrection of interest in Virtual Reality (VR) over recent years has motivated us to investigate the implementation of affective recognition systems within virtual environments, in ongoing attempts to increase immersion and engagement levels. Therefore, as this study tackles this academically challenging issue, we believe that this chapter delivers a valuable contribution to the field of affective computing and emotion recognition.

Chapter 6
Individual Psychophysiological Differences

Abstract – In Chapter 5, by employing the cross-validation technique, 112 classifiers (using KNN and SVM) were trained to classify the physiological database into either Affective Clusters or Emotion Labels, under various pre-processing settings. However, the performances of the trained classifiers need further evaluation within new datasets. To do this, in the present chapter, another physiological experiment, logging the EEG, GSR and heart rate of 15 participants who did not take part in any of the previous experiments, was conducted. The analysis highlighted that the trained classifiers perform either similar to, or worst than random classifiers, within the new dataset. The analysis also demonstrated a significant individual difference between participants, which prevented the classifiers to be generalised within new datasets. Although it was concluded that the presented affective computing systems could be considered as subject-dependent classifiers; it was also highlighted that identification and elimination of the sources of these individual differences should be considered as important as the affective classification process, and is worthy of further investigation.

6. Individual Psychophysiological Differences

6.1. Psychophysiological Evaluation Database Construction (Evaluation Experiment – *Experiment 5*)

6.1.1. Participants and Method

To evaluate the performance of the affective recognition system described in Chapter 5, using new datasets (rather than cross-validation), another physiological experiment has been conducted. In this experiment, the Cosine similarity algorithm (Pang-Ning Tan, 2005) has been employed to identify the most affective sub-game, in each Affective Cluster (as described in Section 3.5.2). Moreover, the neutral game in the Preliminary Experiment (Experiment 3 – Section 3.5.2) has been added to the four most affective sub-games (one sub-game in each Affective Cluster). Fifteen gamer participants were recruited to participate in this experiment (9 males and 6 females – 24 (±5) years old – none had participated in the previous experiments). The same room, equipment and experiment software, employed in the physiological experiment (Section 5.2) were used in this experiment. Firstly, the participants completed a training session (Section 3.5.3) and the neutral sub-game (Section 3.5.2). Then they played the four most affective games in a random order. The EEG, GSR and heart rates of the participants, while playing the sub-games, were recorded and stored in Microsoft Excel files (as described in Chapter 5 – refer to Appendix H). The participants were instructed to self-assess their emotional experiences, at the end of each sub-game, using both dimensional and categorical models (Section 2.3.1). The study was reviewed and approved by the University of Birmingham's Ethical Review Committee (Ethical Reference Number: ERN_13-1157).

6.1.2. Results

The EEG, GSR and heart rates of the participants, during all 75 affective sessions[73], were recorded and tagged by the emotional experience of the participants, self-assessed at the end of each affective session. The filtering, normalisation, windowing, feature extraction and defective data removal steps, as described in Section 5.3, were performed on the raw physiological database, to construct the *evaluation feature matrix*. However, instead of all 743 features (extracted as described in Section 5.3.3), only the unique most optimal features (230 features identified in Section 5.7.2) were extracted in the evaluation feature matrix construction process. This was due to the fact that only the unique most optimal features are employed by the best performing classifiers (Section 5.7.3 and Appendix L), and are required to perform the evaluation process. As the best performing classifiers (Section 5.7.3 and Appendix L) employed the Hamming windowing technique, all 28 Hamming-based windows were employed in the windowing process. Therefore, 28 evaluation feature matrices (according to 28 windowing techniques – Section 5.3.2.1), each of which consisted of all 230 unique most optimal features, were generated. These evaluation feature matrices were used in the affective recognition system assessment process.

[73] $15\ (Participants) \times 5\ (Sub-Games) = 75\ (Affective\ Sessions)$

6.2. Evaluation of the Affective Recognition System

6.2.1. Results

All of the 112 best performing KNN and SVM classifiers[74] (Section 5.7.3 and Appendix L), trained according to the training feature matrices (Section 5.3), have been assessed within the evaluation feature matrices (Section 6.1.2). To do this, the required features for each classifier have been extracted from the corresponding evaluation feature matrix (according to the window length –Refer to Appendices J and L). Then, the classification results of the classifiers (either the identified Affective Cluster or Emotion Label), within the final 70% of a single sub-game (Section 5.3.3.6), have been compared to the self-assessment of the participant, at the end of that particular sub-game. To conduct the performance evaluation, the system's classification accuracy and F1-Score have been measured.

Figure 37 presents the accuracies of the classifiers, with respect to different window lengths. The dashed horizontal lines present the accuracy of random classifiers, categorising the physiological responses into either Affective Clusters[75] or Emotion Labels[76]. On average, the KNN and SVM classifiers achieved (respectively) a 30.43% (±3.58) and 39.16% (±3.45) classification accuracy with respect to the Affective Clusters. However, they achieved considerably lower classification accuracies (KNN=16.92% (±2.66) and SVM 19.25% (±3.02)) when employing the Emotion Labels. Indeed, the accuracies of the classifiers were significantly higher than the corresponding random classification accuracy ($P<0.001$[77]).

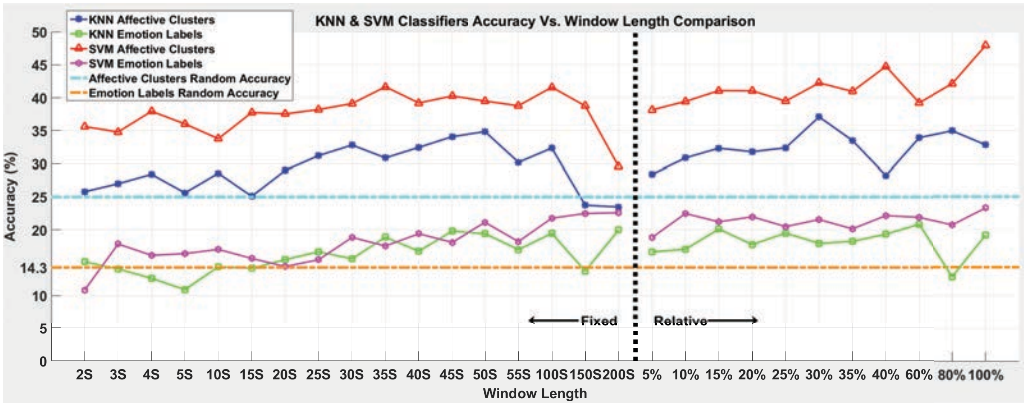


Figure 37 – Accuracy Comparison of KNN and SVM Classifiers, According to Both Affective Clusters and Emotion Labels Within the Evaluation Feature Matrix – The Dashed Vertical Line Separates the Fixed and Relative Windows – The Dashed Horizontal Lines Present the Accuracies of a Random Classifiers (25% and 14.3% for Classification According Affective Clusters and Emotion Labels, Respectively)

Figure 38, on the other hand, presents the performance of the classifiers, according to their mean F1-Scores (across the Affective Clusters or Emotion Labels).

---

[74] 28 for each KNN and SVM classifiers, according to both Affective Clusters and Emotion Labels

[75] As there are 4 classes in Affective Clusters, a random classification is expected to result a 25% ($\frac{1}{4}\times 100 = 25$) accuracy.

[76] As there are 7 classes in Emotion Labels (sadness was not evoked by any game), a random classification is expected to result a 14.3% ($\frac{1}{7}\times 100 = 14.3$) accuracy.

[77] According to a one-sample T-Test, to check if the mean accuracy (across windows) of each classifier is significantly different from the corresponding random classification accuracy.

Moreover, Table 23 presents the mean F1-Scores for Affective Clusters, Emotion Labels and classifiers. The F1-Scores of the KNN classifier, when employing the Affective Clusters, were slightly better than random classification ($P_{Affective\_Clusrter}$ =0.019[78]), but were not significantly different from a random classifier, when employing Emotion Labels ($P_{Emotion\_Labels}$ =0.280[78]). The F1-Scores of the SVM classifier, on the other hand, were significantly worse than random classifiers (P<0.001[78]), with respect to either Affective Clusters or Emotion Labels.

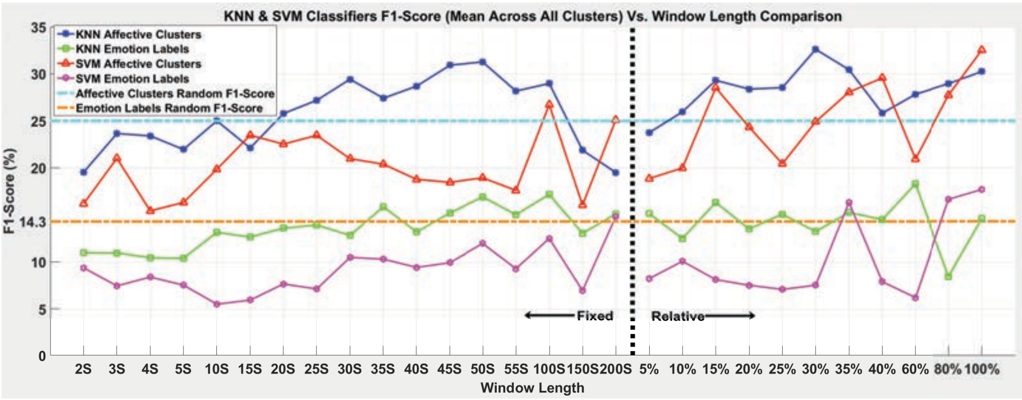


Figure 38 – F1-Score Comparison of KNN and SVM Classifiers, According to Both Affective Clusters and Emotion Labels Within the Evaluation Feature Matrix – The Dashed Vertical Line Separates the Fixed and Relative Windows – The Dashed Horizontal Lines Present the F1-Scores of a Random Classifiers (25% and 14.3% for Classification According Affective Clusters and Emotion Labels, Respectively)

Table 23 – Evaluation Mean F1-Scores Across Classifiers, Classes and All Windowing Techniques – (A - B) Presents the (A) 25th and (B) 75th Percentile

Emotion Label	**Mean F1-Score (25th – 75th Percentiles)**	**Affective Cluster**	**Mean F1-Score (25th – 75th Percentiles)**
Relaxed	5.77% (0.24% - 9.38%)	PVLA	46.90% (40.77% - 50.07%)
Content	14.56% (10.88% - 18.80%)		
Happy	6.09% (0% - 10.19%)	PVHPA	26.02% (18.05% - 35.08%)
Excited	18.35% (11.75% - 24.52%)		
Angry	27.04% (22.08% - 30.75%)	NVPA	16.87% (9.58% - 22.71%)
Afraid	3.53% (0% - 6.59%)		
Sad	Not Available	NVNA	7.63% (0% - 13.63%)
Bored	6.50% (0.12% - 11.58%)		
Emotion Labels Classifier	**Mean F1-Score (25th – 75th Percentiles)**	**Affective Clusters Classifier**	**Mean F1-Score (25th – 75th Percentiles)**
KNN	13.82% (12.73% - 15.71%)	KNN	26.67% (23.70% - 29.14%)
SVM	9.56% (7.48% - 10.38%)	SVM	22.04% (18.82% - 25.01%)

[78] According to a one-sample T-Test, to check if the mean F1-Score (across windows) of each classifier is significantly different from the corresponding random classification F1-Score.

6.2.2. Discussion

As can be seen from the results, the performances of the classifiers, within the new dataset, have been considerably poorer, when compared to their performances according to the cross-validations (compare Table 22 and Table 23). Although the accuracies of the classifiers were significantly higher than random classification accuracies[79], their F1-Scores were, on average, either similar to or worse than random classifiers. To be able to compare the datasets[80], first the training (Experiment 4 – Section 5.3) and evolution (Experiment 5 – Section 6.1.2) feature matrices were combined. Then, 28 subsets of features (for 28 different window lengths) were separated. Each of these 28 subsets contains the training and evaluation feature matrices (containing up to 30 features – refer to Section 5.4), employed by the corresponding KNN and SVM classifiers, for a particular window length (refer to Appendices J and L). Then a Multivariate Analysis of Variance (MANOVA)[81] has been conduced 56 times, each of which within one of the 28 subsets, for both Affective Clusters and Emotion Labels. As multiple significance levels assessments have been conducted, the P-value critical level has been corrected, using the *Bonferroni Correction* algorithm (Hommel, 1983) ($\text{P value critical level} = \frac{0.05}{56} = 0.0008$). The analysis highlighted no significant difference between the training (Experiment 4 – Section 5.3) and evaluation (Experiment 5 – Section 6.1.2) feature matrices, in any of the 28 subsets ($P>0.99$ for all 28 subsets). Therefore one can conclude that this significant classification performance attenuation could be due to significant psychophysiological individual differences, rather than feature matrix variations.

6.3. Individual Differences

6.3.1. Significant Physiological Range Differences

As discussed in Section 3.1, variability between affective experiences, rather than differences among individuals, has been the focus of this study. To minimise the variability between individual emotional experiences (psychological responses), a single controllable Affective VR scenario has been designed, to minimise the background scenario variations (the "Speedboat" simulation) within affective sessions. It was discussed that variable background scenarios could result in extremely inconsistent individual emotional experiences (Section 3.1). On the other hand, as discussed in Section 4.3.3, significant individual differences in emotional experiences among participants with different age, gender and gaming experience have been highlighted. To minimise the individual differences even further, only gamers (from both genders) aged between 18 and 30 were recruited (in both Experiment 4 and 5), as they showed minimum differences in emotional experiences.

[79] 25% and 14.3% when performing random classification according to (respectively) Affective Clusters and Emotion Labels.
[80] In case there is a significant difference between the datasets, recorded in two spate experiments. As an illustration, in case the EEG recordings of Experiment 4 are significantly different from Experiment 5, due to some sort of hardware issues.
[81] The features were considered as the dependent variables, while the **Experiment** (Training *vs*. Evolution), **Subject ID** (a unique code for all 30 and 15 participants in Experiment 4 and 5) and **Emotion Classes** (either Affective Clusters or Emotion Labels) as the independent parameters.

These efforts would, it was proposed, attenuate the individuals' psychological differences, as the sources of affective variations (VR variations, age, gender and gaming experiences) have been adjusted accordingly. However, the physiological responses of the participants could vary with respect to various biological factors (Section 5.1.4).

To analyse these differences, the combined training and evaluation feature matrices (Section 6.2.2) have been employed. A Multivariate Analysis of Variance (MANOVA)[81] analysis highlighted a significant difference between physiological responses of the participants, within different Affective Clusters and Emotion Labels ($P<0.0001$[82] for all 28 subsets – Section 6.2.2). This variation highlights the fact that different sub-games not only resulted in significantly different psychological reactions (Section 4.2), but also caused significantly different physiological responses. From this, one can conclude that the Affective VR designed for this research has been effective in the manipulation of participants' emotional experiences (in terms of both their psychological and physiological responses).

On the other hand, the same Multivariate Analysis of Variance (MANOVA)[81] analysis highlighted a significant difference between different individuals as well ($P<0.0001$[82] for all 28 subsets – Section 6.2.2). Moreover, the intercept of the participants and emotional experiences (Affective Clusters or Emotion Labels) has been identified as an extremely significant factor in the variation of physiological responses ($P<0.0001$[82] for all 28 subsets – Section 6.2.2). These significant factors highlight the fact that, in a single Affective Cluster or Emotion Label, the physiological response of each participant is significantly different from the others.

6.3.2. Significant Psychophysiological Pattern Differences

As discussed above (Sections 6.2.2 and 6.3.1), the trained classifiers performed almost randomly, while detecting the emotional experiences of new participants, due to significant individual differences in psychophysiological behaviours. As the classifiers employ high dimensional feature spaces (17 to 30 features, refer to Appendix L) in the classification process, demonstrating graphical representation of individual differences in the feature spaces is almost impossible. Therefore, two of the most optimal features ("GSR Fluctuation Frequency" and "Maximum Heart Rate"), selected by majority of the classifiers (refer to Table 21), have been selected to create a 2-dimensioanl feature space. Figure 39 presents the 2D feature space scatter plots of four randomly selected participants. As can be seen in the scatter plots, the distribution range and patterns of the features, with respect to the Affective Clusters, are different. Moreover, a MANOVA[83] analysis concluded that the 2D feature space of each individual is different from the others ($P<0.001$). As an illustration, Table 24 presents the participants' distribution range of the GSR fluctuation frequency and maximum heart rate, according to NVNA Affective Cluster. As can be obtained from Table 24 and Figure 39, not only the distribution range, but also the variation pattern

[82] Considering the Bonferroni correction P-value critical value (0.0008).
[83] The features were considered as the dependent variables, while the **Subject ID** and **Affective Clusters** as the independent parameters.

of the GSR fluctuation frequency and maximum heart rate of these four randomly selected participants, when they are labelled within a particular Affective Cluster, are significantly different.

Table 24 – Participants' Distribution Range of the GSR Fluctuation Frequency and Maximum Hear Rate According to NVNA Affective Cluster

Participant	Maximum Heart Rate Range	GSR Fluctuation Frequency Range
1	81 to 102	1000 to 8550
2	88 to 109	280 to 8500
3	99 to 104	370 to 2580
4	68 to 105	210 to 8250

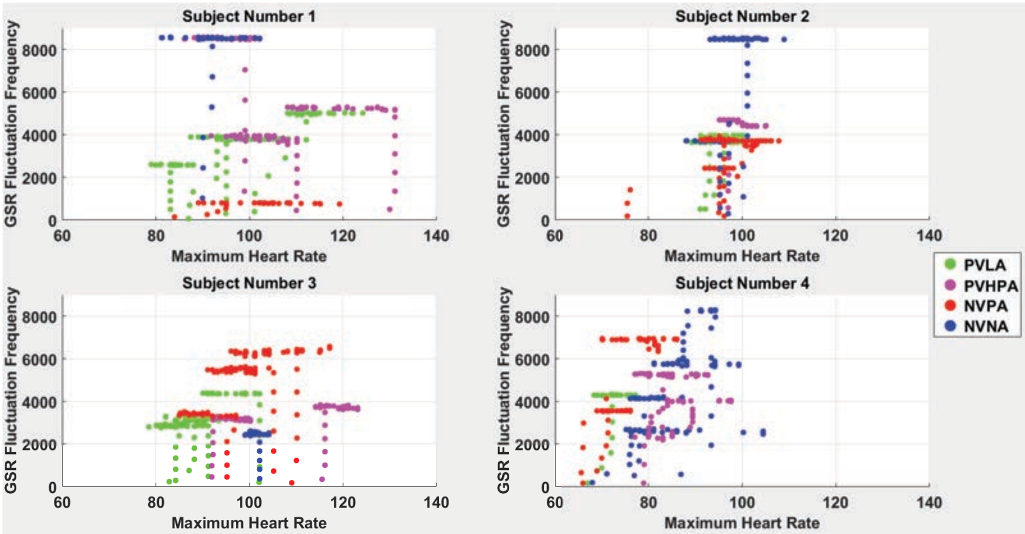


Figure 39 – Raw 2D Feature Space ("GSR Fluctuation Frequency" and "Maximum Hear Rate") Scatter Plot for 4 Randomly Selected Participants – The Colour of the Dots Represents the Affective Cluster, in Which the Dot belongs to

By combining the 2D feature matrices of these four participants, an SVM classifier has been trained to classify their emotional experiences, according to the Affective Clusters. The classifier achieved 77.8% accuracy (F1-Score = 76.3%), according to a 10-fold cross-validation. However, by conducting a leave-one-participant-out validation[84] (to assess the performance of the classifier against a new participant), the classifier achieved 17.16% (F1-Score = 10.33%) accuracy. Then an SVM classifier has been trained on every individual separately to create 4 subject-dependent classifiers, according to the Affective Clusters. Table 25 presents the performance (accuracy and F1-Score) of the four subject-dependent classifiers, according to a 10-fold cross-validation. As can be obtained from the results, the classifiers performed well according to cross-validations, while they performed worse than random classifiers, when new participants are presented to them.

These results present an early demonstration that the physiological pattern of the participants, within different emotional experiences, could vary quite highly.

[84] A SVM classifier has been trained on 3 participants and tested on the remaining one. This process has been repeated 4 times to have every participant as the test subject, at least once.

Therefore, as concluded in Sections 6.2.2 and 6.3.2, and demonstrated in the above example, the psychophysiological range and pattern differences between individuals, prevent any trained classifier from performing better than a random classifier, when new participants are introduced into the prediction process.

Table 25 – Subject-Dependent SVM Classifier Performances According to 2D Feature Space ("GSR Fluctuation Frequency" and "Maximum Hear Rate") – The Presented F1-Score is the Mean of the Classification F1-Scores, Within Each Affective Cluster

Participant	Classification Accuracy	Classification F1-Score
1	80.90%	80.17%
2	83.00%	82.42%
3	70.50%	68.43%
4	78.70%	78.68%

6.4. Normalisation

As discussed in Section 5.1.4, normalisation techniques are employed in highly variable datasets to eliminate participant-related variation ranges. However, as discussed in Section 5.1.4, these techniques require calibration processes, which are highly dependent on their training datasets and could vary dramatically by adding new samples[85]. On the other hand, the normalisation process could be performed on either the raw physiological data or the feature space. Moreover, the parameters of the normalisation technique[85] (Section 5.1.4) could be tuned with respect to either each individual or to the entire space. As an illustration, Figure 40 presents the normalised (according to Z-Score technique – see Section 5.1.4 and Equation 4) 2D feature space scatter plots of the four randomly selected participants (as presented in Section 6.3.2 and Figure 39), according to each individual (subject-based)[86]. Figure 41, on the other hand, presents the normalised (according to Z-Score technique – see Section 5.1.4 and Equation 4) 2D feature space scatter plots of all four participants *combined* [87].

The normalised feature spaces have been employed to train an SVM classifier to classify the emotional experiences of these four participants, according to the Affective Clusters. The classifiers achieved 76.4% (F1-Score = 77.44%) and 77.5% (F1-Score = 78.46%) accuracies, according to the 10-fold cross-validation technique, for subject-based[86] (Figure 40) and combined[87] (Figure 41) normalised datasets, respectively. However, by conducting a leave-one-participant-out validation[84] the classifier achieved an average of 15.73% (F1-Score = 6.9%) and 16.17% (F1-Score = 7.02%) accuracies, for subject-based (Figure 40) and combined (Figure 41) normalised datasets, respectively. Although these results are not sufficient enough to demonstrate any concrete conclusion, they present an early demonstration that the normalisation may not be able to improve the significant classification performance attenuations, in new datasets. This is due to the fact that, although, the normalisation

[85] As an illustration the mean and standard deviation of the heart rate could be changed considerably, if new participants with extremely low or high heart tempos are added to the database (Z-Score Normalisation – Equation 4).

[86] The mean and standard deviation of each participant, separately, have been employed to conduct the Z-Score normalization process.

[87] The mean and standard deviation of all participants, combined, have been employed to conduct the Z-Score normalization process.

techniques can eliminate the significant range differences between the participants (as the normalised data, according to Z-Score technique, ranges between -3 and 3 – refer to Section 5.1.4); the psychophysiological behavioural pattern differences between participants (as presented in Figure 39, Figure 40 and Figure 41) could not be altered.

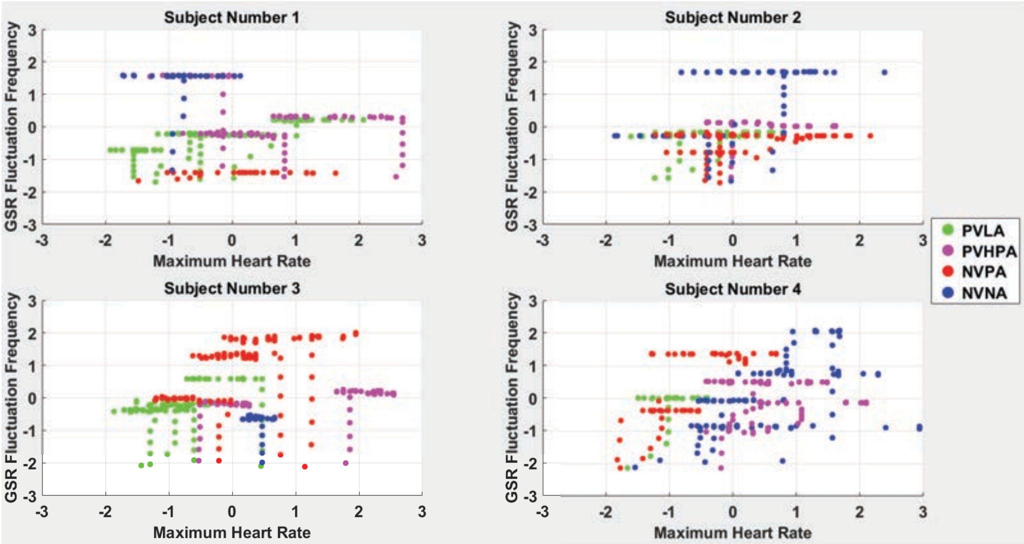


Figure 40 – Subject-Based Normalised 2D Feature Space ("GSR Fluctuation Frequency" and "Maximum Hear Rate") Scatter Plot for 4 Randomly Selected Participants – The Colour of the Dots Represents the Affective Cluster, in Which the Dot belongs to

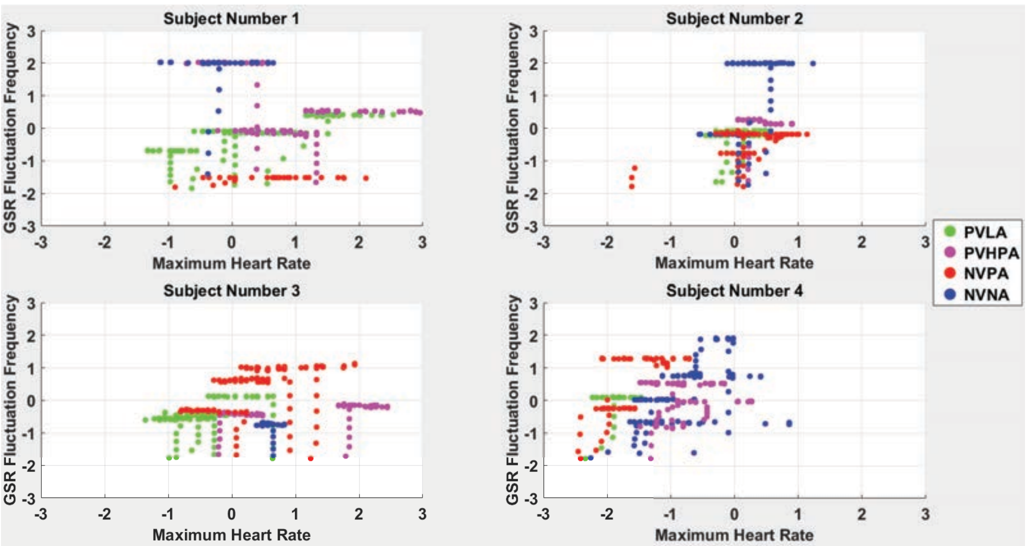


Figure 41 – Normalised (All 4 Participants Combined) 2D Feature Space ("GSR Fluctuation Frequency" and "Maximum Hear Rate") Scatter Plot for 4 Randomly Selected Participants – The Colour of the Dots Represents the Affective Cluster, in Which the Dot belongs to

6.5. Cross-Validation Limitations

As stated in Section 5.6, this study employed a 10-fold cross-validation technique to conduct the classification performance evaluation process. As various windowing techniques (28 windows – Section 5.3.2.4) have been employed in this study, for each affective session (sub-game) there have always been more than two extracted samples in the feature matrix[88] (except the 100% window). Also, as the random folding cross-validation technique has been employed in this study, the classifiers have been able to observe the variation pattern of all participants in the training process. This is due to the fact that, by employing the random folding technique in the cross-validation process, there is an extremely low chance that no window from a particular participant is selected in the training fold, and all extracted windows, from all affective sessions of that particular participant, are used in the evaluation fold[89]. As further evidence for this claim, the cross-validation accuracy of the classifiers is attenuated when a longer window is employed (Section 5.7.3). This can be due to the fact that, in shorter windows, the classifiers have a higher chance of comprehending and memorising larger portions of the participants' variation patterns, and could perform better when assessed against the evaluation fold. Therefore, the classifiers have been able to classify the training feature matrix with high accuracies; while they are highly vulnerable to new datasets with new variation patterns.

As discussed in Section 5.1.6, other cross-validation techniques have been employed in the literature. The K-fold and leave-one-session/observation-out methods show a high level of similarity, as they both employ a large portion of all participants' data for the training process (*K-1* folds in K-fold and *N-1* sessions/observations in leave-one-session/observation-out approach), and the rest for testing purposes. These two techniques have been employed by 72% of the reviewed studies (refer to Section 5.1.6). This is in contrast with the leave-one-participant-out approach, as it employs *N-1* participants for the training process, and all observations from a single participant for testing purposes. This technique has been employed by 28% of the reviewed studies (refer to Section 5.1.6). In the present study, the trained and cross-validated (10-fold) classifiers according to 30 participants (Section 5.7.3 and Appendix L) have been evaluated against 15 new participants (Section 6.2.2). This two-stage validation technique has not been employed by any affective computing study conducted since 1993.

To be able to perform a proper literature comparison, the results of the studies that employed leave-one-participant-out technique have been compared against the results of the present study. This is due to the fact that the leave-one-participant-out cross-validation approach resembles a high level of similarity with the two-stage validation technique conducted in this study (as it employs each individual separately for the testing process). Half of the studies, which employed the leave-one-participant-out cross-validation technique, achieved near or worse than random classification accuracies (Bailenson et al., 2008; Wu et al., 2010; Koelstra et al.,

[88] E.g. 2 80% overlapping windows are extracted from all affective sessions, while around 175 2-second overlapping windows are extracted from a 60 seconds affective session.

[89] Around 3% of the samples within the database contain the physiological responses of a participant within all 10 affective sessions (as there are 30 participants in the training database – Section 5.2). There is a low chance that the randomly selected 10% of the database for the evaluation process contains no window from a particular participant. E.g. considering at least 20 80% windows for each participant. In all affective sessions, there are $\frac{600!}{60!\times(600-60)!}$ possibilities to select 10% of the database for the evaluation fold. While there are $\frac{580!}{40!\times(580-40)!}$ possibilities to select 10% of the database, which does not contain any window for a particular participant, for the evaluation fold. Therefor, in 80% window, there is an almost (3.84×10^{-20})% probability that the training fold contains no window from a particular participant. By decreasing the window length, this chance is exponentially attenuated.

2012). However, the other half that employed this technique and achieved high classification accuracies employed extensive normalisation techniques. They all achieved random classification accuracies, when minimal normalisations were applied on the datasets. They all concluded that, in real case scenarios, where the entire dataset is not available for normalisation parameters' tuning process, the classifiers perform almost randomly when new participants are presented (Soleymani et al., 2012; Wen et al., 2013; Nardelli et al., 2015).

6.6. Robustness to Individual Differences

Turning briefly to experiences in the field of Automatic Speech Recognition (ASR), since the introduction of the ASR in 1950s, various studies have tried to minimise the speaker-dependency of recognition systems (Juang & Rabiner, 2006). The early attempts in the 1950s performed recognition on a limited number of syllables, for a single speaker, and could not be generalised for other individuals (Davis et al., 1952; Olson & Belar, 1956). It was not until the 1970s that ASR systems started to recognise wider ranges of vocabularies, for commercial purposes, with speaker-independency features (Juang & Rabiner, 2006). However, even the most recent systems have demonstrated limited robustness to variability in different aspects, such as: recorder and environmental noise, physical distance of the sound source and the recorder, various accents and grammars, spontaneous-speech phenomenon, individual differences in acoustic characteristics, etc. (Juang & Rabiner, 2006; Omologo et al., 1998). Various filtration, noise cancellation and recognition techniques have been introduced and implemented in the past decades, to improve the performance of ASR systems (with minimum speaker-dependency), to enable them to be reliably employed in commercial applications[90] (Juang & Rabiner, 2006; Omologo et al., 1998).

The affective computing system, designed in the present study, achieved 98.5% maximum cross-validation accuracy. This accuracy is 2% higher than the maximum cross-validation accuracy, achieved by the previous studies reviewed (Section 5.1.5). However, it was concluded that the trained classifiers perform extremely poorly when assessed within new datasets (Section 6.2.2). It has almost been 20 years since the term "Affective Computing" has been introduced into the literature (Picard & Picard, 1997). There are many unexplained physiological individual differences, which considerably minimise the robustness of the designed systems and prevent any further systems generalisation. There is a serious need to identify these sources, to be able to appropriately generalise the classification process. Similar to those endeavours evident within the Automatic Speech Recognition community, the affective computing field requires many further improvements, to be able to introduce robust systems with minimal dependency on individual difference.

6.7. Subject- Independent *vs*. Subject-Dependent Classifiers

In machine learning two types of classifiers can be trained to perform the classification processes: *subject-independent* and *subject-dependent*. Subject-

[90] E.g. Google Now (http://www.google.co.uk/landing/now/) and Apple iOS Siri (http://www.apple.com/uk/ios/siri/).

independent classifiers are designed to analyse variation patterns of datasets to learn and to establish a classification technique to categorise these variation patterns accordingly, regardless of the participants. This means that introducing more or less participants to the dataset should not dramatically alter the classifier performance. In contrast, subject-dependent classifiers perform the same process, but on each individual participant. This means that each classification model is trained according to a specific participant (or a specific set of participants) and can classify his/her (or that particular set of participants') variations, if new observations are recorded. Subject-dependent classifiers are one of the initial enhancement solutions for poor performing classifiers, which do not have any robustness against individual differences. Moreover, it has to be considered that subject-dependent classifiers and learning algorithms could also be considered as reliable machine learning techniques for both academic (Tapia et al., 2007) and commercial[91] purposes. The subject-dependent affective computing technique has been employed by various studies in the past years (Rani et al., 2007; Soleymani et al., 2011; Parnandi et al., 2013; Jenke et al., 2014). Other studies have employed subject-dependent classifiers, either to considerably improve the accuracy of their classifiers (Kim & Andre, 2008), or to enhance their subject-independent classifiers, with random classification accuracies (Bailenson et al., 2008; Wu et al., 2010). Finally, Liu et al. (2009) implemented their subject-dependent classifiers (trained on 15 participants) within a new experiment, with the same participants, to detect their anxiety level in real-time, and to change the game difficulty accordingly. They concluded that, although there was a noticeable drop in the real-time classification accuracy of the classifiers (from 88.9% cross-validation accuracy to 78% real-time accuracy), the overall satisfaction level of the participants was increased considerably, with the emotion-based difficulty adaptation process (Liu et al., 2009).

The initial purpose of the present study was to design a subject-independent affective computing system, capable of classifying participants' emotional experiences through psychophysiological measurements. However, we concluded that the designed classifiers (Section 5.7.3 and Appendix L) are highly dependent on the participants, as the individual psychophysiological differences restrain the generalisation process of the classification models. Therefore it can be concluded that the designed classifiers can be considered as subject-dependent, and could be employed in real-time affective computing applications, to detect the emotional experiences of the same participants that have been recruited in the Primary Experiment (Experiment 4 – Section 5.2). However, it has to be remembered that not only are the classifiers extremely vulnerable to new participants (as demonstrated in Section 6.2.2), they could also be highly susceptible to new emotional patterns. This means that, if the same participants generate different psychophysiological variation patterns (e.g. more intense psychophysiological reaction to stronger emotional experiences, compare to what they have experienced, previously, in the sub-games), the classifiers may not be able to classify their affective experiences accurately.

[91] SwiftKey app that tries to learn the typing habits of a smartphone user, to predict and suggest words, for easier and faster typing experiences: https://swiftkey.com/en/company/

6.8. Conclusion

The original intention of the present study was to design an affective computing system, capable of detecting human emotions within virtual environments, through psychophysiological measurements. During the first stage of this study (Chapter 5), a number of affective computing systems (with different pre-processing and classification settings – refer to Section 5.7.3 and Appendix L), with high cross-validation accuracies, have been designed. In this chapter the affective computing systems have been assessed with new participants, to investigate if the systems have the potential to be implemented within real-time affect recognition applications (Section 6.1 – research question IV presented in Section 1.8). The evaluation process concluded that the trained classifiers perform randomly when new participants are presented (Section 6.2). The analyses highlighted significant individual differences, in both the range and pattern of participants' psychophysiological behaviours (Section 6.3.1). By employing a small portion of the database (two physiological measurements from four randomly selected participants), an early demonstration was presented to discuss the extent of the individual differences (Section 6.3.2). Moreover, two normalisation techniques have been employed as well, to expand this early demonstration, addressing the possibility of improving the classifier performances within new datasets. According to the results, it was concluded that the normalisation of the dataset could not be considered as a reliable solution to enhance the poor performance of the classifiers within new datasets (Section 6.4). By analysing the limitations of the previously reported cross-validation and evaluation techniques, it was highlighted that the vulnerability of the classifiers within new datasets, due to individual differences, had not been appropriately exploited in the literature (Sections 6.5 and 6.6). Finally it was suggested that the designed classifiers could be considered as subject-dependent classifiers and be implemented within affective computing systems, which employ the same participants, in their real-time recognition process (Section 6.7).

The main contribution of this chapter is the design and implementation of the classifier evaluation process, within new datasets (Sections 6.1 and 6.2). As discussed in Section 6.5, the majority of the studies employed cross-validation techniques, to evaluate the performance of their affective recognition systems. This is in contrast with the two-stage evaluation process, implemented within this study, to assess the performance of the designed and cross-validated affective recognition systems, within new datasets. Although, the minority of the affective computing studies employed a leave-one-participant-out cross-validation technique, to simulate the evaluation process within new datasets (as the classifiers are trained on *N-1* participants and tested on the remaining one), they applied extensive normalisation techniques on the entire recorded dataset. Therefore, as the normalisation techniques could not be accurately applied in new datasets or real-time applications (as the normalisation parameters need to be calibrated on the entire dataset), it is believed that conducting this two-stage performance evaluation process has highlighted a significant classification issue, in the area of affective computing and recognition. This contribution has demonstrated a serious need to identify the sources of the individual psychophysiological differences, to be able to appropriately generalise the affective classification process.

Chapter 7

Conclusions and Future Work

7. Conclusions and Future Work

7.1. Conclusions

The human-computer interface has become one of the most important research topics in computer science since the introduction of the first "computers" (calculators) in the 17th century. Today, highly complex real-time computer-based systems and their interfaces with human operators are undergoing an evolution on a hitherto unheard-of scale, in what has become a quest to ensure that they become synergistic, even symbiotic with their human users – transparent, usable, intuitive, sensitive and reactive. As a key part of this evolution, psychophysiological interfaces generally, and Brain-Computer Interaction (BCI) techniques specifically, have introduced new dimensions to the human interaction process, by the introduction of direct human-to-computer connections. Enhancing this symbiosis is, today, both technically and ethically possible. Affective computing systems, as one the sub-classes of BCI research endeavours, have the potential to be exploited in the measurement of users' emotions and affective experiences. Furthermore, by incorporating such systems within advanced simulation-based technologies, such as, for example, Virtual Reality, it may be possible – in the near-term future – to enhance participants' sense of "immersion" and engagement. In the present study, the design, conceptualisation and evaluation of an affective computing system has been demonstrated, implemented in the form of an experimental, dynamic and highly reconfigurable virtual environment.

In Chapter 2 of the thesis, the two most influential models of emotion (Categorical *vs*. Dimensional), representing the affective space, were discussed (research question I presented in Section 1.8). By conducting a 120-participant experiment, the dimensional affective space (Valence, Arousal and Dominance) was categorised into four Affective Clusters, containing eight selected Emotion Labels (Relaxed, Content, Happy, Excited, Angry, Afraid, Sad and Bored). Moreover, by employing the results of the experiments described in Chapters 3, 5 and 6, and presented in two other studies (International Affective Picture System (IAPS) (Lang et al., 2008) and Digital Sound (IADS) (Bradley & Lang, 1999)), it was concluded that the Dominance dimension, introduced by Mehrabian (Mehrabian, 1970), has a high correlation with the Valence axis. Furthermore, it was highlighted that the subjective positioning of the eight Emotion Labels, within the 3-dimensional affective space, only occupies four octants of the space, leaving the other four completely empty ("Negative Valence, Positive Dominance and Positive/Negative Arousal" and "Positive Valence, Negative Dominance and Positive/Negative Arousal"). Therefore, it was concluded that the Circumplex of Affect (Valence *vs*. Arousal), originally introduced by Russell (Russell, 1980), is capable of representing the eight Emotion Labels, adequately, and the Dominance axis could be discarded in the affective assessment process.

In Chapter 3, the development of an adaptive virtual reality system, capable of manipulating the emotional responses of participants, by changing its internal parameters (incidents), was described (research question II presented in Section 1.8).

The Affective VR is capable of generating 792 different variations (sub-games), with potentially different affective powers. By employing the results of a parameter-estimation survey, comprising the responses of 35 participants, coupled with an overall sub-game-approximation algorithm, the affective powers of the sub-games have been estimated. To manipulate the participants' emotions within the entire affective space, the sub-games, which have the highest probability of "pushing" the participants' emotions toward each Affective Cluster, have been identified and selected using the Cosine Similarity algorithm. Consequently, 22 sub-games were presented to 68 participants, with various age, gender and gaming experiences. The most important contribution of this chapter is the design of the fully graphical (3D), dynamic and interactive Affective VR system, capable of evoking multiple emotions on the users. From the literature, it should be remembered that various evaluated affective datasets have been presented in the form of images, sounds, music and video clips, to evoke various emotional experiences. To date, no form of Affective VR has been presented.

In Chapter 4, the results of the experiment conducted in Chapter 3, were analysed (research question II presented in Section 1.8). It was concluded that the designed Affective VR is capable of manipulating the participants' emotional experiences, by changing its internal parameters. Moreover, it was concluded that the emotion forecasting technique is fairly accurate in predicting the colour of the emotional experiences, but underestimated their intensities. However, it was concluded that the dimensional space, when compared to the categorical model, is more reliable in predicting the participants' emotional experiences. On the other hand, it was concluded that the gender, age and gaming experience are significant sources in fluctuating the participants' emotional experiences, when exposed to a single affective stimulus. It was further concluded that participants aged 18 to 24 and 24 to 30, have similar emotional experience patterns when compared to those aged between 30 and 40. Moreover, it was highlighted that the female non-gamers have the least similarity level (in both colour and intensity of the emotion), when compared to the other groups (male gamers, male non-gamers and female gamers). These findings could be considered as the most important contribution of this chapter, as they could provide invaluable insights about the potential emotional response of a user population exposed to a new game (in the gaming industry), prior to its distribution. A capability of being able to identify the potential users (their gender, age and gaming experience) could provide extremely beneficial commercial and marketing information for the diffusion and deployment of the designed games.

In Chapter 5, first a deep literature review, conducted by 30 affective computing studies since 1993, was presented, with the goal of discussing in detail the various stimuli types, physiological measurements, artefact removal and filtering techniques, windowing processes, feature extractions, normalisation algorithms, classification and validation techniques. Then the physiological signal acquisition, pre-processing and feature extraction steps, reviewed in the literature, were implemented in a physiological experiment, consisting of 30 gamer participants (15 male and 15 female, aged between 18 and 30). Following this, 28 feature matrices (developed by employing 28 arbitrary windowing settings), each of which contained 743 features, were extracted from the recorded physiological database. Next, by employing a feature selection technique (minimal-Redundancy-Maximal-Relevance – mRMR), only the 30 *most optimal features* (those, which have the maximum relevance to the classification clusters, with minimum redundancy) were selected. Finally, four classification techniques (KNN, SVM, DA and Classification Tree) were employed to perform the classification process. The performances of more than a quarter of a million classification settings (under various classification and pre-processing settings) were evaluated using the cross-validation technique. The cross-validation results suggested that the KNN (mean classification accuracy 97.01% (±1.28)) and SVM (mean classification accuracy 92.84% (±3.69)) classifiers outperformed the Classification Tree (mean classification accuracy 56.3% (±3.04)) and DA (mean classification accuracy 46.4% (±1.47)) classifiers. Moreover, the cross-validation results suggested that, on average, the classification according to Affective Clusters performed better, when compared to Emotion Labels. The main contribution of this chapter is the employment of all affective computing techniques, which are implemented within different steps of the classification process, by various studies in the past 25 years, to identify the best settings that could generate the best performing classifiers. Another contribution of this chapter is the implementation of affective computing systems within virtual environments (research question III presented in Section 1.8). The resurrection of interest in VR over recent years has provided the motivation for the present research to investigate the implementation of affective recognition systems within virtual environments, in ongoing attempts to increase immersion and engagement levels. Therefore, in tackling this academically challenging issue, it is believed that this chapter delivers a valuable contribution to the field of affective computing and emotion recognition.

In Chapter 6, another physiological experiment, containing 15 new participants, was designed and conducted (9 males and 6 females, all aged between 18 and 30, none participated in previous experiments). The resulting physiological database was employed in another analysis, to evaluate the performance of 112 best performing classifiers (28 classifiers for each windowing settings, for KNN and SVM classifiers, according to both Affective Clusters and Emotion Labels – refer to Appendix L – research question IV presented in Section 1.8). The results suggested that the trained classifiers perform highly accurately when cross-validated with the training dataset, but performed poorly (either worse than, or similar to a random classifier) when tested against a new dataset. Although the results suggested a significant difference between the physiological responses of participants experiencing different emotions, they also highlighted a significant difference amongst individuals as well. It was concluded that, although the Affective VR was able to manipulate participants' emotions and consequently their physiological responses, the unexplained individual differences prevented the classifier training process to be generalised for new datasets. It was concluded that the identification and appropriate adjustment of the sources of these individual differences are extremely important issues for future researchers. Similar to the situation for early automatic speech recognition (ASR) systems, affective computing systems are highly vulnerable to individual differences, and need further improvements to increase their robustness and minimise their individual-dependency. Moreover, it was suggested that affective computing systems could be employed as subject-dependent recognition systems, to recognise the emotional experiences of the same participants, recruited for the training process (Experiment 4 – Section 5.2), in real-time applications. It is believed that conducting this two-stage performance evaluation process has highlighted a significant classification issue, in the area of affective computing and recognition. This is due to the fact that all of the studies in the past 25 years which investigated the affective computing process, employed cross-validation techniques in the evaluation process. This contribution has demonstrated a serious need to identify the sources of the individual psychophysiological differences, to be able to appropriately generalise the affective classification process.

7.2. Future Work

To identify the sources of individual differences, the resting state sessions (i.e. those sessions in which no emotional experience is evoked on the part of the participants) need to be analysed further. There could exist some physiological variation patterns, which are uniquely related to the standard biological variations[92] of a single individual, rather than a particular physiological response of that individual to an emotional experience. By identifying these baseline biological deviations, the physiological variation patterns that are highly related to affective responses can be highlighted and classified. However, the physiological variation patterns, after removing the baseline biological deviations, could also be highly variable between participants, and they may need further generalisation. Nevertheless, it can be concluded that the identification and elimination of individual differences in physiological datasets play a vital role in performing affective classifications.

As discussed in Section 6.7, the affective computing system designed during the present research could be implemented in subject-dependent emotion recognition processes. Therefore, Experiment 5 (Section 6.1) should be repeated with the same participants recruited in Experiment 4 (Section 5.2), to assess the performance of the classifiers, when the same participants in a new experimental setting have been exposed to the Affective VR variations. This future work could clarify the capabilities, robustness and limitations of the subject-dependent emotion recognition systems, designed in this study, in a new affective experience, with the same participants.

Moreover, as another possibility for future research, the possibility of employing regression techniques for designing continuous affective computing systems[93] need to be investigated. The main issue in this investigation is the affective tagging process. As discussed in Section 5.3.3.6, the reported emotional experiences of the participants at the end of each sub-game have been categorised into one of the

[92] E.g. heart rate variations according to a heart characteristic or condition.
[93] In contrast with the designed affective classification systems, which classify the emotional experience of the participants into either Affective Clusters or Emotion Labels.

four Affective Clusters, to be assigned as the average affective label of all windows, extracted from the corresponding sub-game. This process has been conducted by considering the "emotion persistence period", which assumes that the emotional experience of the participants could be variable during a sub-game, but always stays within the categorised Affective Cluster (during the last 70% part of the sub-game – refer to Section 5.3.3.6). To be able to appropriately train regression models, to continuously detect the participants' emotions, another affective tagging technique needs to be considered to assign the 2D emotional experiences (Valence *vs*. Arousal) to each extracted window. The new affective tagging technique has to consider the possible emotional variations of the participants, within a sub-game, as the reported affective experience at the end of each sub-game is the mean emotional experience of the participants, during a particular sub-game.

The final motivation of the present 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 human users' emotions. 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. For example, in command and control for military or counter-insurgency operations, it may be possible to endow multi-input situational awareness display systems with the capability to support end users' decision-making capabilities, generating responses and outcomes based on their instantaneous workload, stress and emotional characteristics, as remote military incidents evolve and crucial tactical and strategic decisions need to be made. Also in the healthcare domain, where the successful use of Virtual Reality in the delivery of real and imaginary scenes to support patients' cognitive restoration or physical/mental rehabilitation depends significantly on their emotional status and their motivation to engage. These are but two applications domains where the complexity of the human perceptual, motor and cognitive subsystems are only now being given the academic and scientific attention they deserve in the future development of symbiotic, engaging and immersive interfaces.

References

Ahlberg, G. et al., 2007. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *The American Journal of Surgery*, 193(6), pp.797–804.

Antje, H. et al., 2005. Emotion studies in HCI-a new approach. In *HCI International Conference*. Las Vegas, 2005.

Bailenson, J.N. et al., 2008. Real-time classification of evoked emotions using facial feature tracking and physiological responses. 66(5), pp.303-17.

Baveye, Y. et al., 2013. A large video data base for computational models of induced emotion. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013*. Geneva, 2013. IEEE Computer Society.

Bergstra, J. & Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13(1), pp.281-305.

Bigand, E. et al., 2005. Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, 19(8), pp.1113-39.

Bradley, M.M., Codispoti, M., Sabatinelli, D. & Lang, P.J., 2001. Emotion and motivation II: Sex differences in picture processing. *Emotion*, 1(3), pp.300-19.

Bradley, M.M. & Lang, P.J., 1994. Measuring Emotion, the Self-Assessment Manikin and the Semantic Differential. *Pergamon Press*, 25(1), pp.49-59.

Bradley, M.M. & Lang, P.J., 1999. *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings*. Tech. Rep. No. B-2. Gainesville: The Center for Research in Psychophysiology, University of Florida.

Bradley, M.M. & Lang, P.J., 2006. Emotion and Motivation. In L.G.T.a.G.B. J.T. Cacioppo, ed. *Handbook of Psychophysiology*. 2nd ed. New York: Cambridge University Press. pp.581-607.

Brown, E. & Cairns, P., 2004. A Grounded Investigation of Game Immersion. In *CHI 'Extended Abstracts on Human Factors in Computing Systems*. Vienna, 2004. ACM.

Cairns, P. et al., 2006. Quantifying the experience of immersion in games. In *workshop on the Cognitive Science of Games and Gameplay at Proceedings of CogSci*. Vancouver, 2006. UCL Interaction Centre - University College London.

Chanel, G., Rebetez, C., Bétrancourt, M. & Pun, T., 2011. Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* , 41(6), pp.1052-63.

Combain, C., D'Argembeau, A., Van der Linden, M. & Aldenhoff, L., 2004. The effect of ageing on the recollection of emotional and neutral pictures. *Memory (Hove, England)*, 12(6), pp.673-84.

Davis, K.H., Biddulph, R. & Balashek, S., 1952. Automatic recognition of spoken digits. *The Journal of the Acoustical Society of America*, 24(6), pp.637-42.

Difede, J. et al., 2007. Virtual Reality Exposure Therapy for the Treatment of Posttraumatic Stress Disorder Following. *The Journal of clinical psychiatry*, 68(11), pp.1639-47.

Ekman, P. & Friesen, W.V., 2003. *Unmasking The Face*. Los Altos: ISHK. ISBN: 0-13-938175-9.

Eugène, F. et al., 2003. he impact of individual differences on the neural circuitry underlying sadness. *NeuroImage*, 19, pp.354-64.

Eysenck, H.J. & Eysenck, M.W., 1985. *Personality and Individual Differences: A Natural Science Approach*. New York: Plenum.

Fernandez-Duque, D. & Landers, J., 2008. Feeling more regret than I would have imagined": Self-report and behavioral evidence. *Judgment and Decision Making*, 3(6), pp.449-56.

Frantzidis, C.A. et al., 2010. 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* , 14(2), pp.309-18.

Frantzidis, C.A. et al., 2010. Toward Emotion Aware Computing: An Integrated Approach Using Multichannel Neurophysiological Recordings and Affective Visual Stimuli. *IEEE Transactions on Information Technology in Biomedicine*, 14(3), pp.589-97.

Gross, J.J. et al., 1997. Emotion and aging: experience, expression, and control. *Psychology and aging*, 12(4), pp.590-9.

Grossman, M. & Wood, W., 1993. Sex Difference in Intensity of Emotional Experience: A Social Role InterPretation. *Journal of Personality and Social Psychology*, 65(5), pp.1010-22.

Hamann, S., 2004. Individual differences in emotion processing. *Current Opinion in Neurobiology*, 14(2), pp.233-38.

Harris, F.j., 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 68(1), pp.51-83.

Hoffman, H.G. et al., 2000. Virtual reality as an adjunctive pain control during burn wound care in adolescent patients. *Pain*, 18(2), pp.305–09.

Hoffman, H.G. et al., 2004. Manipulating presence influences the magnitude of virtual reality analgesia. *Pain*, 11(1-2), pp.162-68.

Hommel, G., 1983. Tests of the overall hypothesis for arbitrary dependence structures. *Biometrical Journal*, 25(5), pp.423–30.

Jack, D. et al., 2001. Virtual Reality-Enhanced Stroke Rehabilitation David. *IEEE Transactions On Neural Systems And Rehabilitation Engineering*, 9(3), pp.308-18.

Jenke, R., Peer, A. & Buss, M., 2014. Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective Computing*, 5(3), pp.327-39.

Joels, M. et al., 2006. Learning under stress: how does it work? *Trends in Cognitive Sciences*, 10(4), pp.152-58.

Juang, B.H. & Rabiner, L.R., 2006. Automatic Speech Recognition – A Brief History of the Technology Development. *Encyclopedia of Language and Linguistics*, 2(1), pp.806–19.

K. Nakajima, T.T.H.M., 1996. Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique. *Medical Engineering & Physics*, 18(5), pp.365–72.

Kaganoff, E., Bordnick, P.S. & Carter, B.L., 2012. Feasibility of Using Virtual Reality to Assess Nicotine Cue Reactivity During Treatment. *Research on Social Work Practice*, 22(2), pp.159-65.

Kantor-martynuska, J. & Horabik, J., 2015. Granularity of Emotional Responses to Music: The Effect of Musical Expertise. *Psychology of Aesthetics, Creativity, and the Arts*, 9(3), pp.235-47.

Katsis, C.D., Katertsidis, N., Ganiatsas, G. & Fotiadis, D.I., 2008. Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach. *IEEE Transactions On Systems, Man, And Cybernetics—Part A: Systems And Humans*, 38(3), pp.502-12.

Kim, J. & Andre, E., 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12), pp.2067-83.

Koelstra, S. et al., 2012. DEAP: a Database for Emotion Analysis Using Physiological Signals. *Affective Computing, IEE Transactions*, 3(1), pp.18-31.

Koelstra, S. & Patras, I., 2013. Fusion of facial expressions and EEG for implicit affective tagging. *Image and Vision Computing*, 31(2), pp.164-74.

Kreyszig , E., Kreyszig , H. & Norminton, E.J., 2010. Mean. Standard Deviation. Variance. Empirical Rule. In *Advanced Engineering Mathematics*. USA: John Wiley & Sons. pp.1013-15.

Kring, A.M. & Gordon, A.H., 1998. Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology*, 74(3), pp.686-703.

Kukolja, D. et al., 2014. Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications. *International Journal of Human Computer Studies*, 72(10-11), pp.717-27.

Lane , D.M. et al., 2013. Transformations. In *Introduction to Statistics*. iTunes eBook. pp.579-81.

Lang, P.J., Bradley, M.M. & Cuthbert, B.N., 2008. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. Technical Report A-8. Gainesville: University of Florida.

Lang, P.J., Gereenwald, M.K., Bradley, M.M. & Hamm, A.O., 1993. Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3), pp.261-73.

Larsen, R.J. & Ketelaar, T., 1991. Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology*, 61(1), pp.132-40.

Leeb, R. et al., 2007. Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: A case study with a tetraplegic. *Computational Intelligence and Neuroscience*, 2007(1).

Lehtinen, M., Forsman, K., Malmivuo, J. & Eskola, H., 1996. Effects of skull and scalp thickness on EEG. *Medical & Biological Engineering & Computing*, 34(1), pp.263-64.

Liu, C., Agrawal, P., Sarkar, N. & Chen, S., 2009. Dynamic Difficulty Adjustment in Computer Games Through Real-Time Anxiety-Based Affective Feedback. *International Journal of Human-Computer Interaction*, 25(6), pp.506-29.

Mahrer, N.E. & Gold, J.I., 2009. The Use of Virtual Reality for Pain Control: A Review. *Current pain and headache reports*, 13(2), pp.100-09.

Mehrabian, A., 1970. A Semantic Space For Nonverbal Behaviour. *Consulting and Clinical Psychology*, 35(2), pp.248-57.

Mohri, M., Rostamizadeh, A. & Talwalkar, A., 2012. *Foundations of Machine Learning*. Massachusetts: MIT Press.

Mroczek, D.K., 2001. Age and Emotion in Adulthood. *Current Directions in Psychological Science*, 10(3), pp.87-90.

Murphy, K.P., 2012. A Simple non-Parametric Classifier: K-Nearest Neighbors. In *Machine Learning A Probabilistic Perspective*. MIT Press. pp.16-18.

Murphy, K.P., 2012. Classification and Regression Trees (CART). In *Machine Learning A Probabilistic Perspective*. MIT Press. pp.546-54.

Murphy, K.P., 2012. Entropy. In *Machine Learning A Probabilistic Perspective*. MIT Press. p.57.

Murphy, K.P., 2012. Estimation the Risk Using Cross Validation. In *Machine Learning: A Probabilistic Perspective*. MIT Press. p.209.

Murphy, K.P., 2012. F-Score. In *Machine Learning: A Probabilistic Perspective*. MIT Press. pp.184-86.

Murphy, K.P., 2012. Gaussian Discriminant Analysis. In *Machine Learning A Probabilistic Perspective*. MIT Press. pp.103-12.

Murphy, K.P., 2012. Introduction. In *Machine Learning: A Probabilistic Perspective*. MIT Press. pp.2-12.

Murphy, K.P., 2012. Kernels. In *Machine Learning A Probabilistic Perspective*. MIT Press. pp.481-88.

Murphy, K.P., 2012. Support Vector Machines. In *Machine Learning A Probabilistic Perspective*. MIT Press. pp.498-507.

Murugappan, M. et al., 2008. Time-Frequency Analysis of EEG Signals for Human Emotion Detection. In *Springer-Verlag*. Berlin, 2008. Springer-Verlag.

Nardelli, M. et al., 2015. Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing*, 6(4), pp.385-94.

Nijholt, A., Plass-Oude Bos, D. & Reuderink, B., 2009. Turning shortcomings into challenges: Brain–computer interfaces for games. *Entertainment Computing 1*, 1(2), pp.85–94.

Novak, D., Mihelj, M. & Munih, M., 2012. A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with Computers 24*, 24(3), pp.154–72.

Olson, H.F. & Belar, H., 1956. Phonetic typewriter. *Acoustical Society of America*, 28(6), pp.1072-81.

Omologo, M., Svaizer, P. & Matassoni, M., 1998. Environmental conditions and acoustic transduction in hands- free speech recognition. *Speech Communication*, 25(1), pp.75-95.

Othman, M. et al., 2013. EEG Emotion Recognition Based on the Dimensional Models of Emotions. *Procedia - Social and Behavioral Sciences*, 97(1), pp.30-37.
Pang-Ning Tan, M.S.V.K., 2005. Introduction to Data Mining. In Pang-Ning Tan, M.S.V.K. *Introduction to Data Mining*. Addison-Wesley. pp.65-73.
Parnandi, A., Son, Y. & Gutierrez-Osuna, R., 2013. A Control-Theoretic Approach to Adaptive Physiological Games. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*., 2013. IEEE.
Parsons, T.D. & Rizzo, A.A., 2008. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3), pp.250-61.
Patrick, E. et al., 2000. Using a Large Projection Screen as an Alternative to Head-Mounted Displays for Virtual Environments. *CHI Letters*, 2(1), pp.478-85.
Peng, H., Long, F. & Ding, C., 2005. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *1226 IEEE Transactions On Pattern Analysis And Machine Intelligence*, 27(8), pp.1226-38.
Picard, R.W. & Picard, R., 1997. Affective Computing. *MIT press*, 252.
Pollak, S.D., Messner, M., Kistler, D.J. & Cohn, J.F., 2009. Development of perceptual expertise in emotion recognition. *Cognition*, 110(2), pp.242-47.
Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P., 1992. Chapter 12 - Fast Fourier Transform. In *Numerical recipes in Fortran (The art of scientific computing)*. Cambridge University Press. pp.490-529.
Rainville, P., Bechara, A., Naqvi, N. & Damasio, A.R., 2006. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, 61(1), pp.5-18.
Rani, P., Sarkar, N. & Adams, J., 2007. Anxiety-based affective communication for implicit human-machine interaction. *Advanced Engineering Informatics*, 21(3), pp.323-34.
Reuderink, B., Mühl, C. & Poel, M., 2013. Valence, arousal and dominance in the EEG during gameplay. *International Journal of Autonomous and Adaptive Communications Systems*, 6(1), pp.45-62.
Rizon, M., Murugappan, M., Nagarajan, R. & Yaacob, S., 2008. Asymmetric Ratio and FCM based Salient Channel Selection for Human Emotion Detection Using EEG. *WSEAS Transactions on Signal Processing*, 4(10), pp.596-603.
Rizzo, A.A. et al., 2002. Virtual environments for the assessment of attention and memory processes: the virtual classroom and office. *Virtual Reality*, pp.3-12.
Rizzo, A.S. et al., 2013. Virtual Reality Applications to Address the Wounds of War. *Psychiatric Annals*, 43, pp.123-38.
Robinson, M.D. & Clore, G.L., 2001. Simulation, Scenarios, and Emotional Appraisal: Testing the Convergence of Real and Imagined Reactions to Emotional Stimuli. *Personality and Social Psychology Bulletin*, 27(11), pp.1520-32.
Rodríguez, A. et al., 2015. Expert Systems with Applications Assessing brain activations associated with emotional regulation during virtual reality mood induction procedures. *Expert Systems with Applications*, 42(3), pp.1699-709.

Russell, J.A., 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6), pp.1161-78.

Salvendy, G., 2006. *Handbook of Human Factors and Ergonomics*. 3rd ed. US: John Wiley & Sons.

Sanei, S. & Chambers, J., 2009. Brain Rhythms. In *EEG Signal Processing*. West Sussex: John Wiley & Sons. pp.10-13.

Sanei, S. & Chambers, J., 2009. Filtering and Denoising. In *EEG Signal Processing*. West Sussex: John Wiley & Sons. pp.79-83.

Schubert, E., 2007. Locus of emotion: The effect of task order and age on emotion perceived and emotion felt in response to music. *Journal of music therapy*, 44(4), pp.344-68.

Seymour, N.E. et al., 2002. Virtual Reality Training Improves Operating Room Performance. *ANNALS OF SURGERY*, 236,(4), pp.458–64.

Silvia, P. & Berg, C., 2011. Finding Movies Interesting: How appraisals and expertise influence the aesthetic experience of film. *Empirical Studies of the Arts*, 29(1), pp.73-88.

Soleymani, M., Asghari-Esfeden, S., Fu, Y. & Pantic, M., 2015. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, 7(1), pp.17 - 28.

Soleymani, M., Koelstra, S., Patras, I. & Pun, T., 2011. 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. IEEE.

Soleymani, M., Pantic, M. & Pun, T., 2012. Multimodal emotion recognition in response to videos. *IEEE Transactions on Affective Computing*, 3(2), pp.211-23.

Stone, R.J. & Allardice, R., 1996. *A Competitive Study on the Virtual Reality Market: Executive Summary*. Prepared & Edited for the Technology Foresight VR Sub-Committee of the UK Department of Trade & Industry (Invitation to Tender No. 01825).

Stone, R.J. et al., 2017. A “Mixed Reality” Simulator Concept for Future Medical Emergency Response Team Training. *Journal of the Royal Army Medical Corps*.

Stone, R.J. & Hannigan, F.P., 2014. Applications of Virtual Environments: An Overview. In Hale, K.S. & Stanney, K.M. *Handbook of Virtual Environments: Design, Implementation and Applications*. CRC Press (Taylor & Francis). pp.883-957.

Sutton, S.K. & Davidson, R.J., 2010. Prefrontal Brain Asymmetry : Inhibition Systems. *Psychological science*, 8(3), pp.204-10.

Takahashi, K. & Tsukaguchi, A., 2003. Remarks on Emotion Recognition from Multi-Modal Bio-Potential Signals. In *Systems, Man and Cybernetics, 2003. IEEE International Conference*. Yamaguchi Univ., Japan, 2003. IEEE.

Tapia, E.M., Intille, S.S. & Haskell, W., 2007. Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor. In *11th IEEE International Symposium on Wearable Computers*. Boston, USA, 2007. IEEE.

Van Kleef, G.A., 2009. How Emotions Regulate Social Life. *Current Directions in Psychological Science*, 18(3), pp.184-89.

Wagner, N., Hassanein, K. & Head, M., 2010. Computer use by older adults: A multi-disciplinary review. *Computers in Human Behavior*, 26(5), pp.870–82.

Wen, W. et al., 2013. Emotion Recognition Based on MultiVariant Correlation of Physiological Signals. *IEEE Transactions on Affective Computing*, 5(2), pp.126-40.

Wild, B., Erb, M. & Bartels, M., 2001. Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: Quality, quantity, time course and gender differences. *Psychiatry Research*, 102(2), pp.109-24.

Wolpaw, J.R. et al., 2002. Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, pp.767–91.

Wu, D. et al., 2010. Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task. *IEEE Transactions on Affective Computing*, 1(2), pp.109-18.

Yazdani, A., Lee, J.-S. & Ebrahimi, T., 2009. Implicit Emotional Tagging of Multimedia Using EEG Signals and Brain Computer Interface. In *The first SIGMM workshop on Social media*. Beijing, 2009. ACM.

Yerkes, R. & Dodson, J., 1908. The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Comparative Neurology and Psychology*, 18(1), pp.459-82.

Zelenski, J.M. & Larsen, R.J., 1999. Susceptibility to affect: a comparison of three personality taxonomies. *Journal of personality*, 67(5), pp.761-91.

Zyda, M., 2005. From visual simulation to virtual reality to games. *Computer*, 38(9), pp.25 - 32.

Appendix A

1. What value of these parameters would describe the experience of " Being Relaxed" in virtual realities?

	-3	-2	-1	0	1	2	3
Valence (Displeasure/pleasure)	☐	☐	☐	☐	☐	☐	☐
Arousal (Sleepy / Aroused)	☐	☐	☐	☐	☐	☐	☐
Dominance (No Control on Game / Full Control on Game)	☐	☐	☐	☐	☐	☐	☐

2. What value of these parameters would describe the experience of " Being Content" in virtual realities?

	-3	-2	-1	0	1	2	3
Valence (Displeasure/pleasure)	☐	☐	☐	☐	☐	☐	☐
Arousal (Sleepy / Aroused)	☐	☐	☐	☐	☐	☐	☐
Dominance (No Control on Game / Full Control on Game)	☐	☐	☐	☐	☐	☐	☐

3. What value of these parameters would describe the experience of " Being Happy" in virtual realities?

	-3	-2	-1	0	1	2	3
Valence (Displeasure/pleasure)	☐	☐	☐	☐	☐	☐	☐
Arousal (Sleepy / Aroused)	☐	☐	☐	☐	☐	☐	☐
Dominance (No Control on Game / Full Control on Game)	☐	☐	☐	☐	☐	☐	☐

4. What value of these parameters would describe the experience of "Being Excited" in virtual realities?

	-3	-2	-1	0	1	2	3
Valence (Displeasure/pleasure)	☐	☐	☐	☐	☐	☐	☐
Arousal (Sleepy / Aroused)	☐	☐	☐	☐	☐	☐	☐
Dominance (No Control on Game / Full Control on Game)	☐	☐	☐	☐	☐	☐	☐

5. What value of these parameters would describe the experience of "Being Afraid" in virtual realities?

	-3	-2	-1	0	1	2	3
Valence (Displeasure/pleasure)	☐	☐	☐	☐	☐	☐	☐
Arousal (Sleepy / Aroused)	☐	☐	☐	☐	☐	☐	☐
Dominance (No Control on Game / Full Control on Game)	☐	☐	☐	☐	☐	☐	☐

6. What value of these parameters would describe the experience of "Being Angry/Annoyed" in virtual realities?

	-3	-2	-1	0	1	2	3
Valence (Displeasure/pleasure)	☐	☐	☐	☐	☐	☐	☐
Arousal (Sleepy / Aroused)	☐	☐	☐	☐	☐	☐	☐
Dominance (No Control on Game / Full Control on Game)	☐	☐	☐	☐	☐	☐	☐

7. What value of these parameters would describe the experience of "Being Sad" in virtual realities?

	-3	-2	-1	0	1	2	3
Valence (Displeasure/pleasure)	☐	☐	☐	☐	☐	☐	☐
Arousal (Sleepy / Aroused)	☐	☐	☐	☐	☐	☐	☐
Dominance (No Control on Game / Full Control on Game)	☐	☐	☐	☐	☐	☐	☐

8. What value of these parameters would describe the experience of "Being Bored" in virtual realities?

	-3	-2	-1	0	1	2	3
Valence (Displeasure/pleasure)	☐	☐	☐	☐	☐	☐	☐
Arousal (Sleepy / Aroused)	☐	☐	☐	☐	☐	☐	☐
Dominance (No Control on Game / Full Control on Game)	☐	☐	☐	☐	☐	☐	☐

Appendix B

Emotion Labels Ratings Distribution:

Figure 1 to Figure 8 Show the distribution of the Emotion Labels ratings, within the 3-dimensional affective space. The red dashed rectangle presents the borders of the Affective Cluster, which contain the corresponding Emotion Labels. However, the blue dashed rectangles present the other Affective Clusters. The purple dots represent the distribution of the Emotion Labels ratings, while the dot sizes present the frequency percentage of that particular rating, among the subjective ratings.

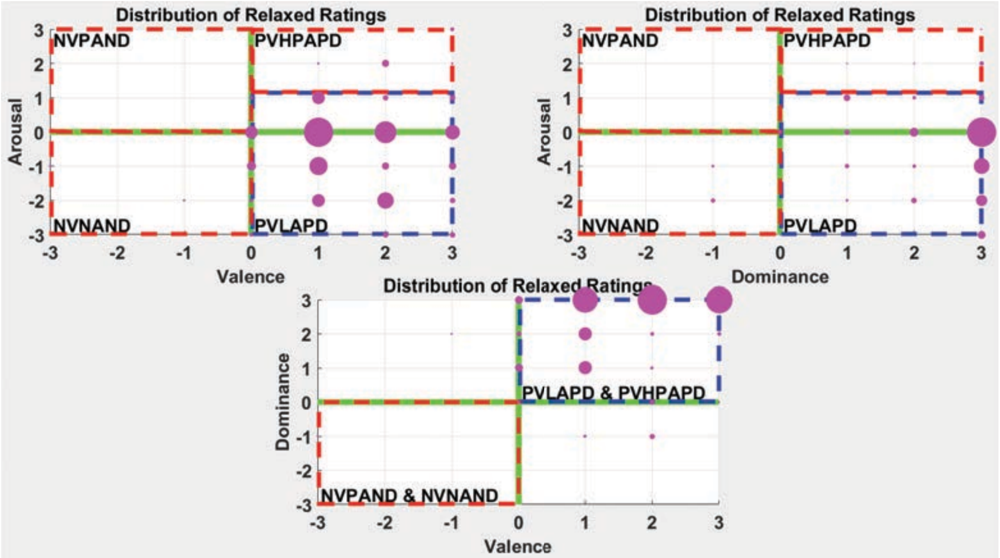


Figure 1 – The Rating Distribution of Relaxed Label

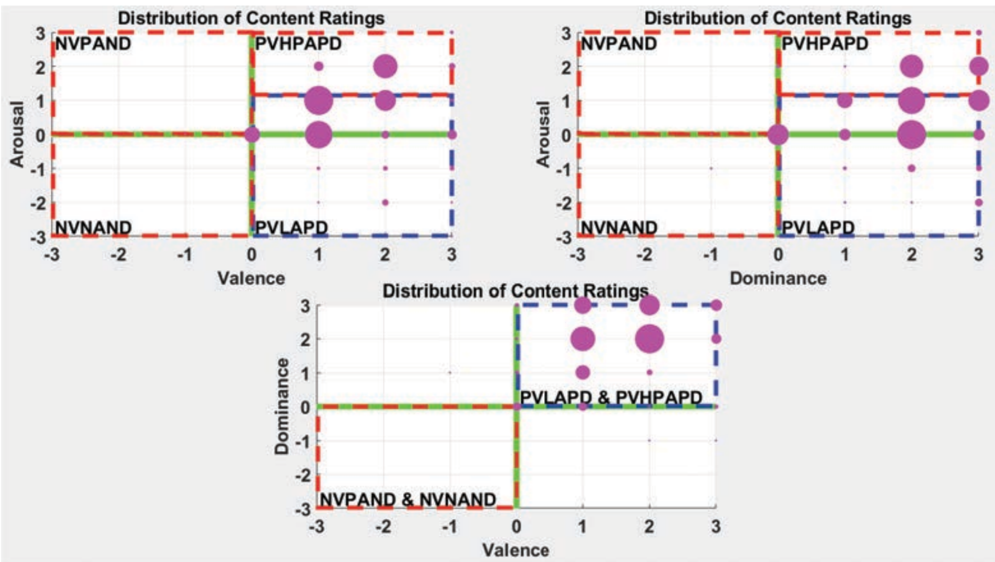


Figure 2 – The Rating Distribution of Content Label

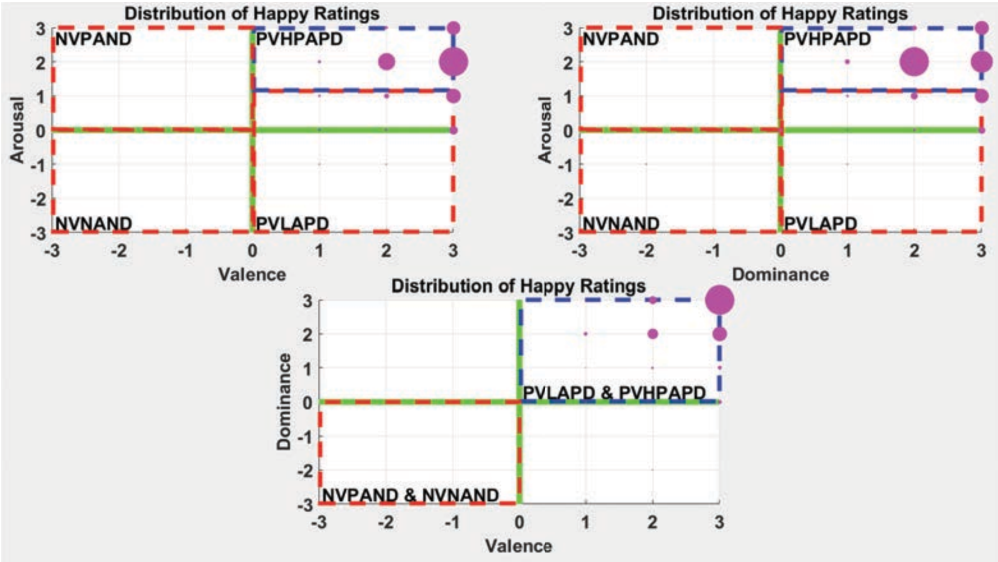


Figure 3 – The Rating Distribution of Happy Label

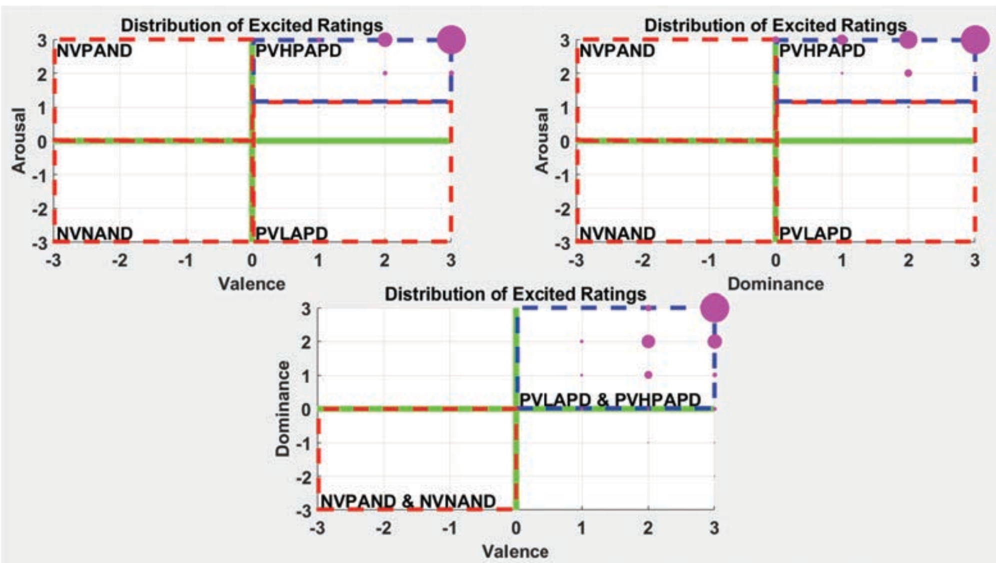


Figure 4 – The Rating Distribution of Excited Label

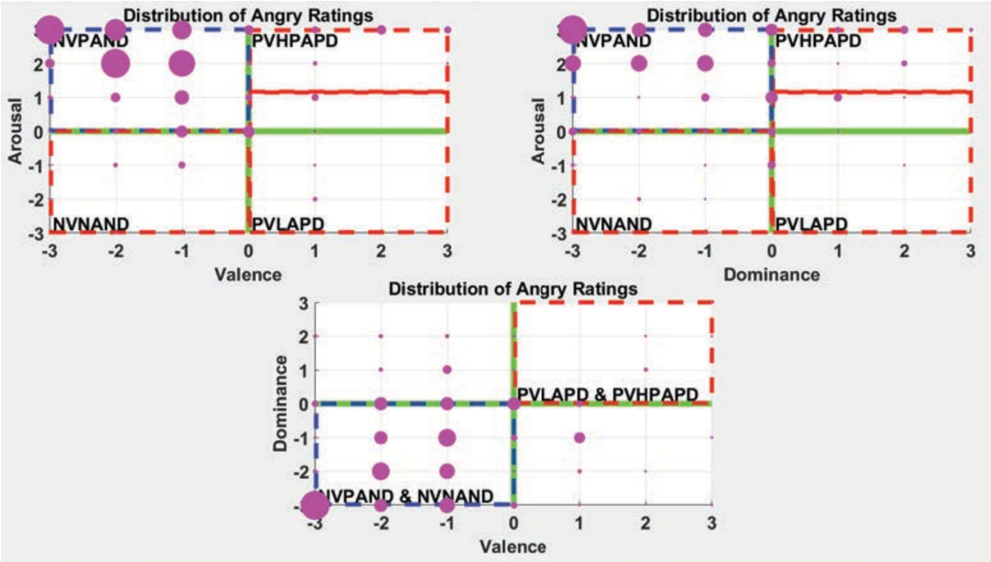


Figure 5 – The Rating Distribution of Angry Label

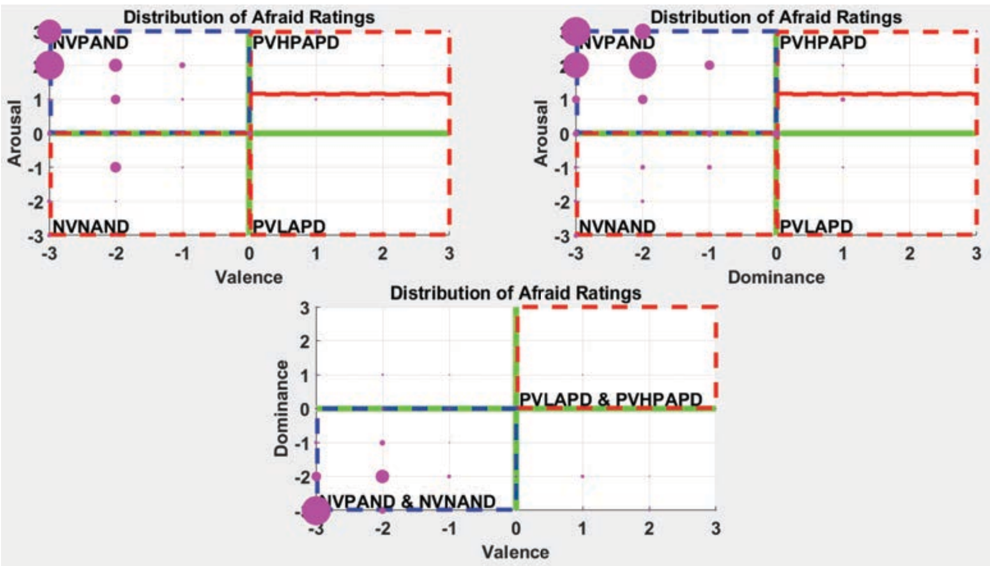


Figure 6 – The Rating Distribution of Afraid Label

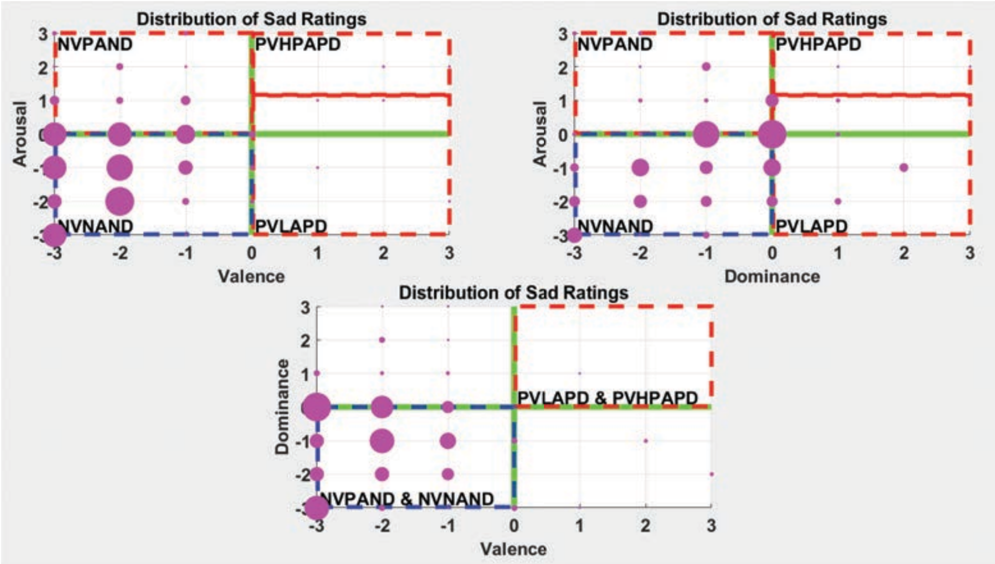


Figure 7 – The Rating Distribution of Sad Label

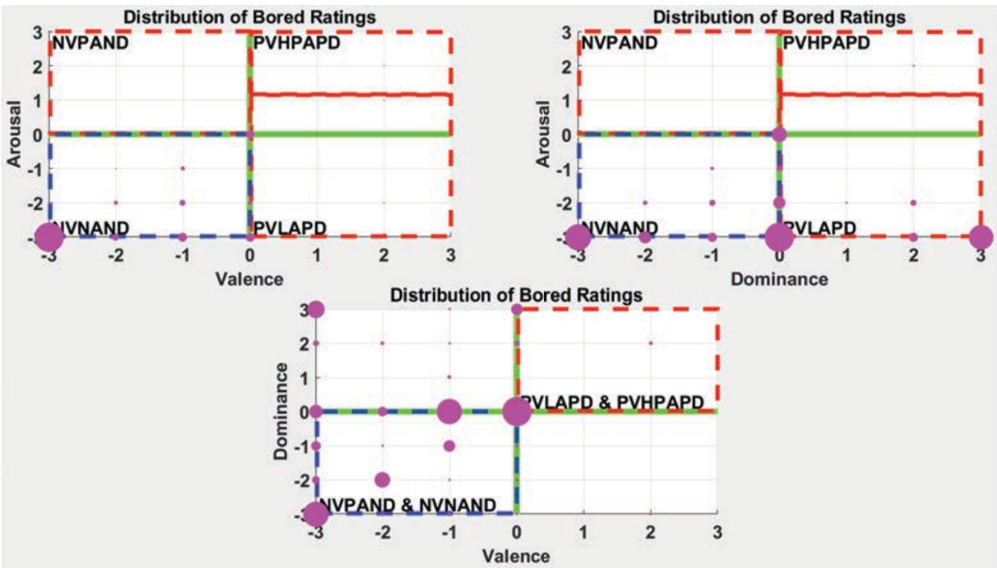


Figure 8 – The Rating Distribution of Bored Label

Appendix C

Report Abuse

Page 1 / 9

The Experiment Explanation

This experiment is designed to assess the emotional impact of different game manipulations. You would be presented with 16 game manipulations. Please rate each on 3 scales:
1) Valence: How pleasurable this game experience would be. Higher positive value means more pleasure (you like it) and higher negative value means displeasure (you dislike it).
2) Arousal: How arousing this game experience would be. Higher positive value means more aroused (e.g. excited, alert, stressful, etc.) and higher negative value means negatively aroused (e.g. sleepy, tired, bored, etc.).
3) Dominance: How much control you have on the game. Higher positive value means more control on the game and higher negative value means the game is out of your control.
In addition, we ask you to choose the most appropriate emotion that fits the game experience.

1* **Please state if you are a gamer or not (if you follow games in the market regularly and have a lot of experience playing games on PC and consoles you are a gamer)?**

- I am a gamer
- I am NOT a gamer

2* **Gender**

- Male
- Female

3* **Age**

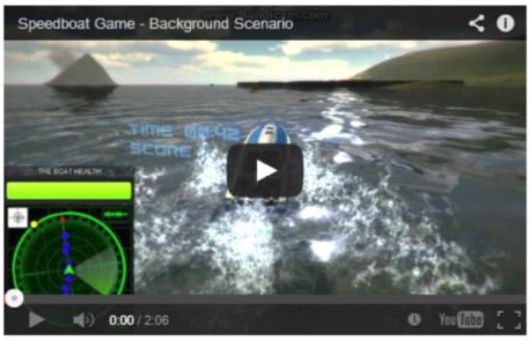
- 12 to 18
- 18 to 24
- 24 to 30
- 30 to 40
- 40 to 50
- 50 to 60
- above 60

Next Page

14 January 2014
21:21

Overall Game and Background Scenario Explanation

The game is a speedboat simulator. You must rescue people from a sinking ship by collecting floats while driving towards it. The more floats you collect, the more people you can rescue. You control the boat via mouse. A blue arrow points to the nearest float. A boat-radar shows the positions of the sinking ship, your speedboat and the floats (mines and torpedoes in case of their presence). A health bar above the radar indicates the remaining health condition of the boat. The game is over when there is no health remaining. Please watch the video of the basic scenario below.
Assuming the basic scenario explained above, consider each of the following incidents individually. Try not to mix incidents from previous questions with the incident in the current question you are answering.



4 * **Do you like this game or not**

- ○ Like
- ○ Unlike

Previous Page | Next Page

Screen clipping taken: 23/02/2014 15:50

Task and Contextual Related Incidents

6* **Playing the game as it was described earlier with no additional features.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

7*

- ○ Relaxed
- ○ Content
- ○ Happy
- ○ Excited
- ○ Angry/Annoyed
- ○ Afraid
- ○ Sad

- Bored

8* **There is no sinking ship. You are allowed to drive freely and explore the environment and collect floats.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

9*

- Relaxed
- Content
- Happy
- Excited
- Angry/Annoyed
- Afraid
- Sad
- Bored

10* **There are several explosive mines on the way that need to be avoided. In case of any collision the health of the boat would be decreased.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence							

(Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

11*

- ○ Relaxed
- ○ Content
- ○ Happy
- ○ Excited
- ○ Angry/Annoyed
- ○ Afraid
- ○ Sad
- ○ Bored

12* **Multiple torpedoes are fired at you from different locations simultaneously. Sound, arrow and additional visual cues provide information of their location. In case of any hit the health of your boat would be decreased. The boat can be manoeuvred to avoid the torpedoes.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance							

(No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

13*

- ○ Relaxed
- ○ Content
- ○ Happy
- ○ Excited
- ○ Angry/Annoyed
- ○ Afraid
- ○ Sad
- ○ Bored

14* **Aside from collecting floats and driving towards the target, there will be a flying ball in front of the boat. Every time you hit it with a water splash gun, you gain task relevant rewards. The water splash gun is on the boat and is controlled by an additional different controller (means you have to use two controllers at the same time).**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

15*

- Relaxed
- Content
- Happy
- Excited
- Angry/Annoyed
- Afraid
- Sad
- Bored

16* **Instead of free open sea, you drive in a complex and long maze.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

17*

- Relaxed
- Content
- Happy
- Excited
- Angry/Annoyed
- Afraid
- Sad
- Bored

Game-Setting and Environmental Related Incidents

18* **There is a time limitation for reaching the sinking boat and there is a penalty associated if you late.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

19*

- Relaxed
- Content
- Happy
- Excited
- Angry/Annoyed
- Afraid

- Sad
- Bored

20* **There is a time limitation for reaching the sinking boat and there is a penalty associated if you late. However, the timer is faulty and randomly increment the seconds too fast.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

21*

- Relaxed
- Content
- Happy
- Excited
- Angry/Annoyed
- Afraid
- Sad
- Bored

22* **The water contains some invisible barriers that you do not see. If you hit them there would be an explosion and you would lose some of your health.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	◯	◯	◯	◯	◯	◯	◯
Arousal (Sleepy / Aroused)	◯	◯	◯	◯	◯	◯	◯
Dominance (No Control on Game / Full Control on Game)	◯	◯	◯	◯	◯	◯	◯

23*

- ◯ Relaxed
- ◯ Content
- ◯ Happy
- ◯ Excited
- ◯ Angry/Annoyed
- ◯ Afraid
- ◯ Sad
- ◯ Bored

24* **You have to control the boat in stormy weather condition.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	◯	◯	◯	◯	◯	◯	◯
Arousal (Sleepy / Aroused)	◯	◯	◯	◯	◯	◯	◯
Dominance (No Control on Game / Full Control on Game)	◯	◯	◯	◯	◯	◯	◯

25*

- Relaxed
- Content
- Happy
- Excited
- Angry/Annoyed
- Afraid
- Sad
- Bored

Previous Page | Next Page

Page 5 / 9

Interaction Related Incidents

26* **In comparison to using a mouse you use joystick to control the boat. The slider on the joystick accurately changes the speed of the boat and also allows backwards movement. This would make me feel:**

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

27*

- ○ Relaxed
- ○ Content
- ○ Happy
- ○ Excited
- ○ Angry/Annoyed
- ○ Afraid

- Sad
- Bored

28* **In comparison to using a mouse you use a *force feedback* joystick to control the boat. The slider on the joystick accurately changes the speed of the boat and also allows backwards movement. The joystick vibrates based on the speed of the boat and the stick pushes to right and left based on the waves hitting the boat.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

29*

- Relaxed
- Content
- Happy
- Excited
- Angry/Annoyed
- Afraid
- Sad
- Bored

30* **The controller of the game (either mouse or joystick) is faulty (e.g. reaction lag, inverted directions, momentarily stops**

working, etc.).

This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

31 *

- ○ Relaxed
- ○ Content
- ○ Happy
- ○ Excited
- ○ Angry/Annoyed
- ○ Afraid
- ○ Sad
- ○ Bored

Previous Page | Next Page

Visualisation Related Incidents

32* **The whole game is to be played in black and white mode (no colours, compare it with the colourful mode). This would make me feel:**

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

33*

- ○ Relaxed
- ○ Content
- ○ Happy
- ○ Excited
- ○ Angry/Annoyed
- ○ Afraid
- ○ Sad

- Bored

34* **The whole game is to be played in inverse black and white mode (no colours, like the negative films, compare it with the colourful mode).**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

35*

- Relaxed
- Content
- Happy
- Excited
- Angry/Annoyed
- Afraid
- Sad
- Bored

36* **The camera would shake and become slightly blur in case of any jump, collision or other similar accidents.**
This would make me feel:

	-3	-2	-1	Zero	1	2	3
Valence							

(Displeasure/pleasure)	◯	◯	◯	◯	◯	◯	◯
Arousal (Sleepy / Aroused)	◯	◯	◯	◯	◯	◯	◯
Dominance (No Control on Game / Full Control on Game)	◯	◯	◯	◯	◯	◯	◯

37*

- ◯ Relaxed
- ◯ Content
- ◯ Happy
- ◯ Excited
- ◯ Angry/Annoyed
- ◯ Afraid
- ◯ Sad
- ◯ Bored

Previous Page | Next Page

Defining Emotions

Based on your previous game experiences, please rate the three parameters (valence, arousal and dominance) to describe the following 8 different emotions.

38* What value of these parameters would describe the experience of " Being **Relaxed**" in virtual realities?

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

39* What value of these parameters would describe the experience of " Being **Content**" in virtual realities?

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○

Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

40* What value of these parameters would describe the experience of " Being **Happy**" in virtual realities?

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

41* What value of these parameters would describe the experience of "Being **Excited**" in virtual realities?

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

42* What value of these parameters would describe the experience of

"Being **Afraid**" in virtual realities?

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

43* What value of these parameters would describe the experience of "Being **Angry/Annoyed**" in virtual realities?

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

44* What value of these parameters would describe the experience of "Being **Sad**" in virtual realities?

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○

Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

45* What value of these parameters would describe the experience of "Being **Bored**" in virtual realities?

	-3	-2	-1	Zero	1	2	3
Valence (Displeasure/pleasure)	○	○	○	○	○	○	○
Arousal (Sleepy / Aroused)	○	○	○	○	○	○	○
Dominance (No Control on Game / Full Control on Game)	○	○	○	○	○	○	○

Previous Page | Next Page

Appendix D

Cosine Similarity:

Cosine similarity is a measure of similarity between 2 vectors in an n-dimensional space (Figure 1). The cosine of the angle between two vectors compares the directions in which each vector is pointed out. As an illustration, if the vectors were in the same direction, the $Cos(\propto)$ would be maximised ($Cos(\propto) = 1$); whereas if they are in complete opposite directions, the $Cos(\propto)$ will be minimised ($Cos(\propto) = -1$). Figure 2 presents 3 similarity level comparisons between two 2-dimensional vectors. In an n-dimensional space the cosine similarity could be calculated using Equation 1, while A and B represent the vectors.

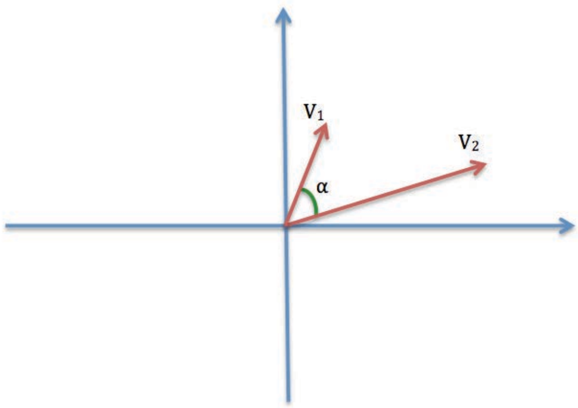


Figure 1 – Two 2-dimensional Vectors – α Represents the Angle Between These 2 Vectors

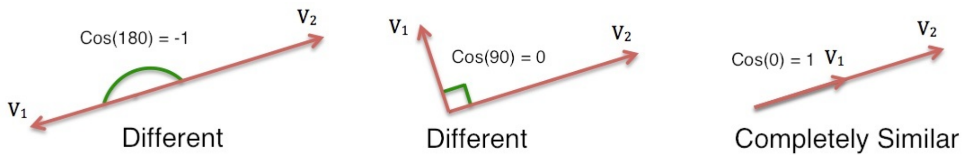


Figure 2 – Cosine Similarity For α=0, α=90 and α=180

$$Similarity\ Level = Cos(\propto) = \frac{A \, . \, B}{|A| \times |B|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Equation 1 – Cosine Similarity Formula in n-Dimensional Spaces – A_i and B_i Represent the i^{th} Elements of Vector A and B, Respectively

Appendix E

Consent Form

Date	
Start Time	
Finish Time	

1. Please state if you are a gamer or not (if you follow games in the market regularly and have a lot of experience playing games on PC and consoles you are a gamer)?
   - ☐ **I am a gamer**
   - ☐ **I am not a gamer**

2. Gender
   - ☐ **Male**
   - ☐ **Female**

3. Age
   - ☐ **12 to 18**
   - ☐ **18 to 24**
   - ☐ **24 to 30**
   - ☐ **30 to 40**

Please confirm the followings:

- ☐ I have read the information sheet and I am fully aware of the purpose and procedure of the experiment.
- ☐ I am participating in this experiment voluntary and I am fully aware that I can leave the experiment at any time.
- ☐ I know that my data will be completely anonymous and can be deleted from the data base on my demand in the future.

Date: ………………………………………

Sign: ………………………………………..

Appendix F

Experiment Purpose:

This experiment is designed to assess the emotional impact of different games. You will be presented with 22 games, each of which last between 2-5 minutes and has its own characteristics and objectives. You are asked to fill in a questioner after each game to assess your emotional experience during the game.

Experiment:

The background scenario for all of the games will stay the same. In the beginning of the experiment you will pass a training session. The objectives, skills and environment will be introduced during the training session. The game score will be calculated at the end of each game. You are supposed to gain the highest possible score for each game.

Although each game can have its own tasks and objectives, all games can be **finished** by passing through the **"Finish Line"** flag. In games that have time limitation, you **have to** pass through the "Start Line" flag to start the timer, otherwise the game score will be calculated as zero. Also if you lose all of the **boat health**, the game will be finished with zero score. Moreover, If you do not manage to finish the game in the limited time provide, the game will be finished with zero score. If you spend more than 5 minutes in the games with **no time limitation** (means you do not pass the finish line by then), the game will be automatically ended, and your score will be calculated.

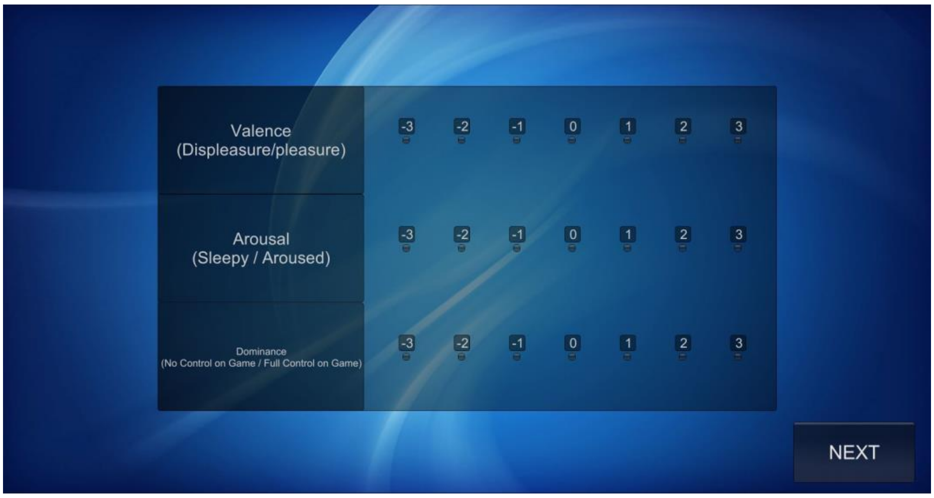
Before each game, an introduction page will introduce the overall and important game's tasks, objectives and controller. Please **consider** each of them carefully; otherwise you will **miss** the objective of the game. Also during some games some or all parts of the game may behave incorrectly:

1. **Faulty Timer**: The timer can be faulty and randomly increase the time.
2. **Faulty Controller**: The controller also can stop responding, or even behave wrongly (ex. Reverse the controller axis).
3. **Invisible Barrier**: There may be invisible barriers in the game with stop signs on them. In case of getting too close to them, they will turn visible and reduce the boat health in case of any collision.

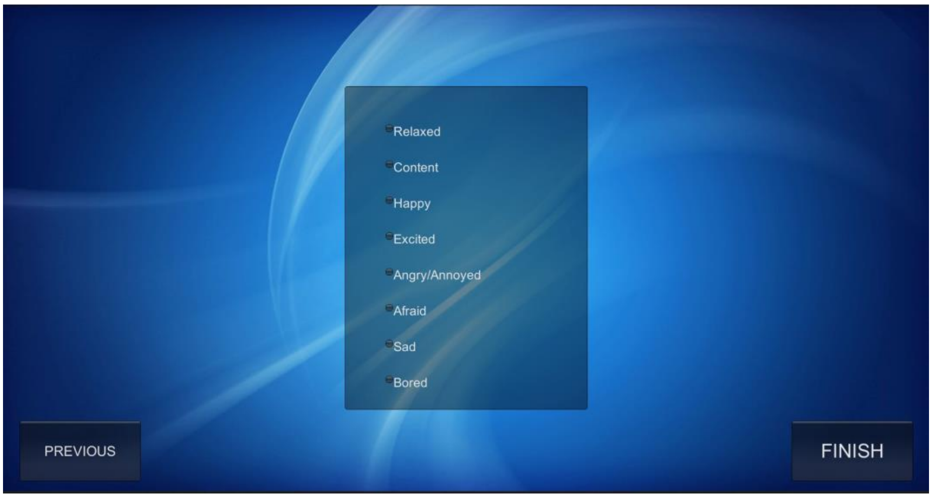
Questioner:

At the end of each game you are asked to rate your emotional experience during the game. The ratings are based on 3 scales:

1. **Valence**: How pleasurable this game experience was. Higher positive value means more pleasure (you like it) and higher negative value means displeasure (you dislike it).
2. **Arousal**: How arousing this game experience was. Higher positive value means more aroused (e.g. excited, alert, stressful, etc.) and higher negative value means negatively aroused (e.g. relaxed, sleepy, tired, bored, etc.).
3. **Dominance**: How much control you have on the game. Higher positive value means more control on the game and higher negative value means the game is out of your control.



In addition, we ask you to choose the most appropriate emotion label, from a list, that fits the game experience.



Please try to assess your emotional experience **during** the game, and try **not** to answer the questioner based on the outcome of the game.

Experiment Duration:

The experiment will last at most 100 minutes. After each questioner, you can rest as long as you prefer and start the next game as soon as seems appropriate. Although, we will appreciate if you finish the experiment completely, you are allowed to leave the experiment after each questioner at any stage.

Appendix G

Paper Index	Year	Article Name
1	2012	Multimodal emotion recognition in response to videos
2	2015	Analysis of EEG signals and facial expressions for continuous emotion detection
3	2011	Continuous emotion detection in response to music videos
4	2010	Prefrontal Brain Asymmetry : Inhibition Systems
5	2015	Assessing brain activations associated with emotional regulation during virtual reality mood induction procedures
6	2013	A control-theoretic approach to adaptive physiological games
7	2008	Asymmetric ratio and FCM based salient channel selection for human emotion detection using EEG
8	2008	Time-Frequency Analysis of EEG Signals for Human Emotion Detection
9	2003	Remarks on emotion recognition from multi-modal bio-potential signals
10	2010	Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task
11	2013	Valence, arousal and dominance in the EEG during game play
12	2010	On the Classification of Emotional Biosignals Evoked While Viewing Affective Pictures: An Integrated Data-Mining-Based Approach for Healthcare Applications
13	2005	Emotion Studies in HCI – a New Approach A new approach to structuring and representing emotions in an HCI system
14	2008	Toward Emotion Recognition in Car-Racing Drivers: A Biosignal Processing Approach
15	1993	Looking at pictures: affective, facial, visceral, and behavioral reactions.
16	2011	Emotion assessment from physiological signals for adaptation of game difficulty
17	2009	Dynamic Difficulty Adjustment in Computer Games Through Real-Time Anxiety-Based Affective Feedback
18	2007	Anxiety-based affective communication for implicit human-machine interaction
19	2006	Basic emotions are associated with distinct patterns of cardiorespiratory activity
20	2008	Emotion recognition based on physiological changes in music listening
21	2008	Real-time classification of evoked emotions using facial feature tracking and physiological responses
22	2012	DEAP : A Database for Emotion Analysis Using Physiological Signals
23	2014	Feature extraction and selection for emotion recognition from EEG
24	2013	Fusion of facial expressions and EEG for implicit affective tagging
25	2015	Recognizing emotions induced by affective sounds through heart rate variability
26	2014	Comparative analysis of emotion estimation methods based on physiological measurements for real-time applications
27	2013	Emotion Recognition Based on Multivariate Correlation of Physiological Signals
28	2013	EEG Emotion Recognition Based on the Dimensional Models of Emotions
29	2010	Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli
30	2009	Implicit emotional tagging of multimedia using EEG signals and brain computer interface

Table 1 – Literature Titles, Year of Publication and Reference Index

Paper Index	# Subjects	Assessment Technique	Emotion Assessment	Physiological Signals								Stimuli					Normalisation			Affective Analyse (Correlation, Statistical Analysis, etc.)	Classification or Regression Technique												Validation Technique	Classification Accuracy
				EEG (# Channels)	GSR	ECG HR	Blood Pressure	Pupil Diameter	EMG	Breathing	Skin Temperature	Video	Music/ Sound	Image	Real Life Scenario	VR	Raw Signals	Features	Technique		Random Forests	Quadratic Bayes Normal (QBN)	Gaussian Naïve Bayes (GNB)	K Nearest Neighbours	Support Vector Machine	Linear Regression	Non-Linear Regression	Fuzzy C-Mean	Neural Network	Decision Tree	Discriminant Analysis	Fuzzy Logic		
1	20	Self Ratings	Dimensional Model	✓ (32)	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓	Min-Max	✗	✗	✗	✗	✗	✓ (RBF Kernel)	✗	✗	✗	✗	✗	✗	✗	Cross	76.40%
2	28	Self Ratings	Dimensional Model	✓ (32)	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓	Z-Score	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	Cross	NA
3	32	Self Ratings	Dimensional Model	✓ (32)	✓	✓	✓	✗	✓	✓	✓	✓	✗	✗	✗	✗		✗		✗	✗	✗	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗	Cross	NA
4	46	Self Ratings	BIS/BAS Personality	✓ (29)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	Log-Transformation	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	NA
5	24	Self Ratings	5 Emotion Labels	✓ (EMOTIV)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓		✗		✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	NA
6	20	Hypothesis (GSR Fluctuations positive Correlation with Arousal)	Arousal Level	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓		✗		✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	NA
7	5	Self Ratings	5 Emotion Labels	✓ (63)	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	Z-Score	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	NA	NA
8	5	Self Ratings	5 Emotion Labels	✓ (63)	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	Z-Score	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	NA	NA
9	?	Hypothesis (Stimuli with Assumed Valance Level)	Pleasure/Displeasure	✓ (2)	✗	✓	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗	✓	Z-Score	✗	✗	✗	✗	✗	✓ (Gaussian and 1st-4th order polynomial Kernel)	✗	✗	✗	✓	✗	✗	✗	Cross	59.7 & 62.3%
10	19	Hypothesis (Estimating Arousal based on Performance)	Arousal Level	✓ (7)	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓	✗	✓	Min-Max	✗	✗	✗	✗	✗	✓ (Linear Kernel)	✗	✗	✗	✗	✓	✗	✗	Cross	96.50%
11	12	Self Ratings	Dimensional Model	✓ (32)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓		✗		✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	NA
12	28	IAPS Ratings	Dimensional Model	✓ (19)	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗		✗		✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	Cross	77.68%
13	31	Self Ratings	Dimensional Model	✗	✓	✓	✗	✓	✗	✗	✓	✗	✗	✓	✓	✗		✗		✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	NA
14	10	Emotion Expert	4 Emotion Labels	✗	✓	✓	✗	✗	✓	✓	✗	✗	✗	✗	✓	✗		✗		✗	✗	✗	✗	✗	✓ (RBF Kernel)	✗	✗	✗	✗	✗	✗	✗	Cross	78.00%
15	66	Self Ratings	Dimensional Model	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗	✓	✗	✗	✓	✗	Log-Transformation	✗	✗	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	Cross	NA
16	20	Self Ratings	3 Emotion Labels	✓ (19)	✓	✗	✓	✗	✗	✓	✓	✗	✗	✗	✗	✓		✗		✗	✗	✗	✗	✗	✓ (RBF Kernel)	✗	✗	✗	✗	✗	✓	✗	Cross	63.00%
17	15	Self Ratings	3 Emotion Labels	✗	✓	✓	✗	✗	✓	✗	✓	✗	✗	✗	✗	✓	✓	✗	Min-Max	✗	✗	✗	✗	✓	✓	✗	✓	✗	✗	✓	✗	✗	Cross	88.50%
18	5	Self Ratings	Anxiety Range 1-9	✗	✓	✓	✗	✗	✓	✗	✓	✗	✗	✗	✓	✗		✗		✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓	Cross	95.00%
19	43	Self Ratings	4 Emotion Labels	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✓	✗		✗		✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	Cross	65.30%
20	3	Self Ratings	Dimensional Model	✗	✓	✓	✗	✗	✓	✓	✗	✗	✓	✗	✗	✗	✓	✗	Z-Score	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	Cross	95.00%
21	41	Hypothesis (Using Face Recognition Software)	3 Emotion Labels	✗	✓	✓	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗		✗		✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	Cross	79.00%
22	32	Self Ratings	Dimensional Model	✓ (32)	✓	✓	✗	✗	✓	✓	✓	✗	✓	✗	✗	✗		✗		✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	Cross	62.00%
23	16	IAPS Ratings	Dimensional Model	✓ (64)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✓	Z-Score	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	Cross	NA
24	24	Self Ratings	Dimensional Model	✓ (32)	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗		✗		✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	Cross	69.00%
25	27	IAPS Ratings	Dimensional Model	✗	✗	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗		✗		✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	Cross	83.00%
26	14	Self Ratings	5 Emotion Labels	✗	✓	✓	✗	✗	✗	✓	✓	✗	✗	✓	✗	✗	✓	✗	Min-Max	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	Cross	60.30%
27	101	Self Ratings	5 Emotion Labels	✗	✓	✓	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	Z-Score	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	Cross	74.00%
28	5	Hypothesis (Stimuli with Assumed Valance/Arousal)	Dimensional Model	✓ (2)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗		✗		✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	NA	NA
29	28	IAPS Ratings	Dimensional Model	✓ (19)	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗		✗		✗	✗	✗	✗	✗	✓ (Gaussian and 1st-4th order polynomial Kernel)	✗	✗	✗	✗	✗	✗	✗	Cross	81.30%
30	8	Self Ratings	6 Emotion Labels	✓ (32)	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗	Min-Max	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	Cross	80.19%

Table 2 – Overall Studies Review

Paper Index	EEG Filter			HR Filter		GSR Filter	
	Artefact Removal	Lower Band (Hz)	Upper Band (Hz)	Lower Band (Hz)	Upper Band (Hz)	Lower Band (Hz)	Upper Band (Hz)
1	Minimum Movement Instruction	4	45	NA	NA	NA	NA
2	Minimum Movement Instruction	4	45	NA	NA	NA	NA
3	Minimum Movement Instruction + Eye Artefact Removal Using EOG	4	45	None Applied	None Applied	None Applied	None Applied
4	Deleting Periods with High Artefact	0	200	NA	NA	NA	NA
5	Minimum Movement Instruction + Eye Artefact Removal Using EOG	5	45	NA	NA	NA	NA
6	NA	NA	NA	NA	NA	None Applied	None Applied
7	None	0.05	45	NA	NA	NA	NA
8	None	0.05	70	NA	NA	NA	NA
9	None	4	45	None Applied	None Applied	None Applied	None Applied
10	Eye and Muscle Artefact Removal Using EOG	1	30	None Applied	None Applied	None Applied	None Applied
11	Eye and Muscle Artefact Removal Using EOG	0.2	128	NA	NA	NA	NA
12	Adaptive Filter Based on The LMSs Algorithm	0.5	40	None Applied	None Applied	None Applied	None Applied
13	NA	NA	NA	None Applied	None Applied	None Applied	None Applied
14	NA	NA	NA	100	500	Moving Average	Moving Average
15	NA	NA	NA	None Applied	None Applied	None Applied	None Applied
16	Visually Checked	4	45	None Applied	None Applied	Moving Average	Moving Average
17	NA	NA	NA	None Applied	None Applied	None Applied	None Applied
18	NA	NA	NA	None Applied	None Applied	None Applied	None Applied
19	NA	NA	NA	None Applied	None Applied	None Applied	None Applied
20	NA	NA	NA	None Applied	None Applied	0	0.2
21	NA	NA	NA	None Applied	None Applied	None Applied	None Applied
22	Eye Artefacts Removal Using Blind Source Separation Technique	4	45	None Applied	None Applied	None Applied	None Applied
23	None	0.1	100	NA	NA	NA	NA
24	None	4	45	NA	NA	NA	NA
25	NA	NA	NA	None Applied	None Applied	NA	NA
26	NA	NA	NA	None Applied	None Applied	None Applied	None Applied
27	NA	NA	NA	0.5	35	0	10
28	None	None Applied	None Applied	NA	NA	NA	NA
29	Adaptive Filter Based on The LMSs Algorithm	0.5	40	NA	NA	NA	NA
30	Minimum Movement Instruction + Eye Artefact Removal Using EOG	1	12	NA	NA	NA	NA

Table 3 – Physiological Signal Filtering

Paper Index	Stimuli Type	Stimuli Duration (s)	EEG			GSR			Heart Rate		
			Type	Duration (s)	Overlap	Type	Duration (s)	Overlap	Type	Duration (s)	Overlap
1	Video	120	Rectangular	15	50%	NA	NA	NA	NA	NA	NA
2	Video	35-117	Rectangular	1	50%	NA	NA	NA	NA	NA	NA
3	Music Video	60	Rectangular	65	0%	Rectangular	65	0%	Rectangular	65	0%
4	Unknown	60	Hamming	60	50%	NA	NA	NA	NA	NA	NA
5	VR	150	Rectangular	150	0%	NA	NA	NA	NA	NA	NA
6	VR	350	NA	NA	NA	Rectangular	30	0%	NA	NA	NA
7	Video	180	NA	180	NA	NA	NA	NA	NA	NA	NA
8	Video	180	NA	180	NA	NA	NA	NA	NA	NA	NA
9	Music	60	NA	60	NA	NA	NA	NA	Rectangular	60	0%
10	VR	Unknown	Rectangular	3	0%	NA	NA	NA	Rectangular	3	0%
11	Game	120	Rectangular	120	0%	NA	NA	NA	NA	NA	NA
12	IAPS	1	Unknown	2.5	0%	Unknown	2.5	0%	NA	NA	NA
13	Real Life Task	Unknown	NA	NA	NA	Rectangular	Unknown	0%	Rectangular	Unknown	0%
14	Real Life Task	180-240	NA	NA	NA	Rectangular	10	0%	Rectangular	10	0%
15	IAPS	6	NA	NA	NA	Rectangular	10	0%	Rectangular	10	0%
16	Game	300	Rectangular	100	0%	Rectangular	100	0%	NA	NA	NA
17	Game	180	NA	NA	NA	Rectangular	180	0%	Rectangular	180	0%
18	Real Life Task	Unknown	NA	NA	NA	Rectangular	Unknown	0%	Rectangular	Unknown	0%
19	Real Life Task	90	NA	NA	NA	Rectangular	90	0%	Rectangular	90	0%
20	Music	240	NA	NA	NA	Rectangular	160	0%	Rectangular	160	0%
21	Video	120	NA	NA	NA	Rectangular	120	0%	Rectangular	120	0%
22	Music Video	60	Rectangular	65	0%	Rectangular	65	0%	Rectangular	65	0%
23	IAPS	4	Hamming	4	0%	NA	NA	NA	NA	NA	NA
24	Video	35-117	Rectangular	35-117	0%	NA	NA	NA	NA	NA	NA
25	IADS	210-330	NA	NA	NA	NA	NA	NA	Rectangular	88	0%
26	IAPS	15	NA	NA	NA	Rectangular	10	0%	Rectangular	10	0%
27	Video	181-591	NA	NA	NA	Rectangular	3.6	0%	Rectangular	18	0%
28	Image	60	Rectangular	60	0%	NA	NA	NA	NA	NA	NA
29	IAPS	1	Rectangular	2.5	0%	NA	NA	NA	NA	NA	NA
30	Video	60	Rectangular	1	0%	NA	NA	NA	NA	NA	NA

Table 4 – Windowing Duration and Techniques

Paper Index	Single Electrodes Overall Spectral Power					Pair-wise (Difference in Spectral Power) Electrodes Overall Spectral Power					Pair-wise (Difference in Spectral Power) Electrodes - Asymmetric Power Density $\frac{P_{left}-P_{right}}{P_{left}+P_{right}}$					Pair-wise (Difference in Spectral Variance) Electrodes - Asymmetric Power Density $\frac{V_{left}-V_{right}}{V_{left}+V_{right}}$					Others		
	Theta	Slow Alpha	Alpha	Beta	Gamma	Theta	Slow Alpha	Alpha	Beta	Gamma	Theta	Slow Alpha	Alpha	Beta	Gamma	Theta	Slow Alpha	Alpha	Beta	Gamma	Mean of the EEG Signal Subtracted by the Mean of the Resting-State Mode	Event Related Oscillatory Activities	$EEG_{w} = \log(\frac{\sum P_i}{\sum (P_i + \alpha)})$
1	✓	✓	✓	✓	✓	✓	✖	✓	✓	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
2	✓	✓	✓	✓	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
3	✓	✓	✓	✓	✓	✓	✖	✓	✓	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
4	✖	✖	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
5	✓	✖	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
7	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✓	✖	✖	✖	✖	✓	✖	✖	✖	✖	✖
8	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✓	✖	✖	✖	✖	✓	✖	✖	✖	✖	✖
9	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✓	✖	✖
10	✓	✖	✓	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
11	✖	✖	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
12	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✓	✖
16	✓	✖	✓	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✓
22	✓	✓	✓	✓	✓	✓	✖	✓	✓	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
23	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
24	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
28	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
29	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✓	✖
30	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✓	✖

Table 5 – Extracted EEG Features

Paper Index	Mean	Min	Max	Standard deviation	Mean of First Derivative	Mean of the Negative Values in First Derivative	Mean of the Positive Values in First Derivative	Mean of Second Derivative	$\frac{\text{Mean of the Negative Values in the First Derivative}}{\text{All Values in the First Derivative}}$	Spectral Power in 0Hz to 2.4Hz Range	Mean on the Peak Values	Number of Picks	conductivity difference among the peak point and the initiation point	[illegible]
3	✓	✗	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓	✗	✗
6	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
10	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
12	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
13	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
14	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
15	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
16	✓	✗	✗	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗
17	✓	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
18	✓	✗	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
19	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
20	✓	✗	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗
21	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
22	✓	✗	✗	✗	✓	✓	✗	✗	✓	✓	✓	✓	✗	✗
26	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
27	✓		✓	✓	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗

Table 6 – Extracted GSR Features

Paper Index	Mean	Standard Deviation	(Mean After Stimuli) - (Mean Before Stimuli)	Mean of the Peaks	Mean of The First Derivative	Spectral Power			
						low Frequency (0.01Hz - 0.04Hz)	Medium Frequency (0.04Hz - 0.15Hz)	High Frequency (0.15Hz - 0.5Hz)	$\frac{Medium\ Frequency\ Power}{High\ Frequency\ Power}$
3	✓	✓	✖	✖	✖	✓	✓	✓	✓
9	✓	✖	✖	✖	✖	✖	✖	✖	✖
10	✓	✖	✖	✖	✖	✖	✖	✖	✖
13	✓	✖	✖	✖	✖	✖	✖	✖	✖
14	✓	✖	✖	✖	✖	✖	✖	✖	✖
15	✖	✖	✓	✖	✖	✖	✖	✖	✖
17	✓	✖	✖	✓	✖	✖	✖	✖	✖
18	✓	✓	✖	✖	✖	✖	✖	✖	✖
19	✓	✓	✖	✖	✖	✖	✖	✖	✖
20	✓	✖	✖	✖	✓	✓	✓	✓	✖
21	✓	✖	✖	✖	✖	✓	✓	✓	✖
22	✓	✓	✖	✖	✖	✓	✓	✓	✓
25	✓	✓	✖	✖	✖	✓	✓	✓	✓
26	✖	✓	✖	✖	✖	✓	✓	✓	✓
27	✓	✓	✖	✖	✓	✖	✖	✖	✖

Table 7 – Extracted Heart Rate Features

Appendix H

Physiological Experiments Technical Report (Software and Hardware)

1. Physiological Recording Equipment

1.1. Physiological Signal Recording Devices Identification:

To perform the physiological signals recording process (EEG, GSR and heart rate), a number of recording kits need to be identified and assessed. Table 1, Table 2 and Table 3 show the available devices and their capabilities. To be able to implement the devices in the experiments and work automatically with the VR functions, availability of Software Development Kits (SDK – preferably in C#, for optimum performance as the VR codes have been implemented in C#), was one of the most important required features. Moreover, to increase the experiment comfort, availability of wireless connection capabilities had been considered as another important feature, as well. According to the recording kits' capabilities, design and accuracy, the EMOTIV EPOC headset, to record EEG, and Shimmer+, to record GSR and heart rate, have been selected. A summary of the recording kits is presented in Sections **Error! Reference source not found.** and 1.1.2.

Table 1 – EEG Recording Kits Comparison

Device Name	Number of Channels	Connection Protocol	SDK Availability	SDK Language	Additional Measurements	Device Wearing Format
OCZ Neural Impulse Actuator[1]	3	USB	NO	NONE	NONE	Headset
NeuroSky[2]	4	Bluetooth	YES	C	NONE	Headset
EMOTIV EPOC[3]	16	Bluetooth	YES	C#	2-Axes Gyroscope	Headset
ENOBIO[4]	8,20 and 32	Bluetooth	YES	C	3-Axes Accelerometer	Cap

[1] https://www.amazon.co.uk/OCZ-NIA-Neural-Impulse-Actuator/dp/B00168VU4U
[2] http://neurosky.com/biosensors/eeg-sensor/biosensors/
[3] http://emotiv.com/
[4] http://www.neuroelectrics.com/products/enobio/

Table 2 – GSR Recording Kits Comparison

Device Name	Connection Protocol	SDK Availability	SDK Language	Additional Measurements	Device Wearing Format
Shimmer+[5]	Bluetooth	YES	C#	3-Axes Accelerometer 3-Axes Gyroscope 3-Axes Magnetometer	Wristband
NeuLog[6]	USB/ Bluetooth	NO	NONE	NONE	Wristband
eSence[7]	USB	NO	NONE	NONE	Wristband

Table 3 – Heart Rate Recording Kits Comparison

Device Name	Connection Protocol	SDK Availability	SDK Language	Additional Measurements	Device Wearing Format
Neology[6]	USB/ Bluetooth	NO	NONE	NONE	Wristband
Scosche Rhythm HRM [8]	Bluetooth	NO	NONE	NONE	Wristband
Nordictrack[9]	Bluetooth	NO	NONE	NONE	Wristband
Shimmer+[5]	Bluetooth	YES	C#	3-Axes Accelerometer 3-Axes Gyroscope 3-Axes Magnetometer	Wristband

1.1.1. EMOTIV EPOC

To record EEG signals, the EMOTIV EPOC[3] recording kit has been selected. This EEG recording kit provides 16 electrodes/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) based on the EEG international 10-20 system. The kit is designed in form of a headset and can be placed easily on the head. The locations of the electrodes have been pre-set on the headset, although it usually requires some small adjustments after placing it on the head. The electrodes are considered as dry electrodes. Meaning that they do not require any conductive paste to adjust electrodes conductance. However, the foams on the electrodes demand lens-washing solution to keep the electrodes conductance level, acceptably high.

The kit provides a 14-bit Analogue to Digital Convertor (ADC) chipset, with 0.51µv accuracy, and 128 Hz sampling rate. The hardware has a number of integrated filters to remove some environmental noises (50 Hz, 60 Hz, etc.). It also provides a wireless connection with the computer and transmits 128-sample packages for all electrodes, every second. The system employs an internal timer to provide time stamps for each package of data. The kit provides SDK packages to enable the users

5 http://www.shimmersensing.com/
6 https://www.fishersci.com/us/en/brand/n/neulog.html
7 https://www.mindfield.de/en/Biofeedback/Products/Mindfield%C2%AE-eSense-Skin-Response.html
8 http://www.scosche.com/rhythm-plus-heart-rate-monitor-armband
9 https://www.nordictrack.co.uk/accessories.html

to implement customised functions (in C#) to accomplish the required purposes. Also the headset provides a 2-axes gyroscope, as well, to determine any head movements or rotations. Figure 1 shows the EMOTIV EPOC headset and how it is placed on the head. Figure 2 shows the location of the EMOTIV EPOC electrodes (channels) on the head, according to the international 10-20 system.

Figure 1 – EMOTIV EPOC[3]

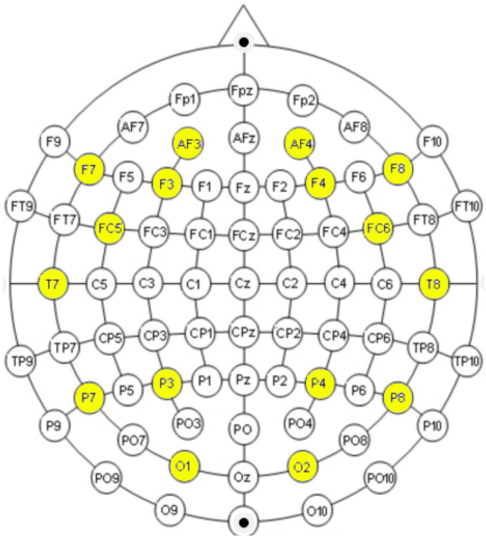


Figure 2 – EMOTIV EPOC Electrodes Locations, according to the International 10-20 System

1.1.2. Shimmer+

To record heart rate and GSR signals, the Shimmer+ kit[5] has been selected. This device provides a wireless platform to record GSR (by 2 finger straps) and PPG (by a light sensor on a finger strap). Figure 3 shows the Shimmer+ kit attached to the right hand. Shimmer+ kit provides a recording platform with adjustable sampling frequency and a 16-bit Analogue to Digital Convertor (ADC). The device has an internal 3-axes gyroscope, accelerometer and magnetometer sensors, as well. The kit provides SDK package to enable the users to develop customised interfaces (in C#), according to their requirements. The system provides internal system time stamps for each package of data. The heart rate can also be extracted from the PPG raw by performing some signal analysis (K. Nakajima, 1996).

Figure 3 – Shimmer+ Kit with the GSR and PPG sensors connected

1.2. Recordable Game Events Identification

It was decided to record a number of game events, as well as the physiological signals. In the following Sections 1.2.1, 1.2.2, 1.2.3 and 1.2.4, *Controllable Incident* is referred to VR incidents (Discussed in Chapter 3); while *Normal Incident* is referred to other normal game events that can happen during any game play (such as game sounds, scores, etc.).

1.2.1. Narrative

During the game play, the user can notice the Time Limitation and Normal/Faulty Timer controllable incidents, as the game time is presented on the game screen. The Current Game Score (presented on the game screen) would show whether the user has gained any score or achievements. Moreover, the Boat Health (presented on the game screen) presents the amount of sustained damages on the boat (by colliding to any mine or torpedo). These 3 game events could be considered as narrative related events, as the importance of task completion timing, game lost and achievements are defined within the context of the game rather than, visualisation, auditory or interactive aspect of the game. Therefore the **Presented Time, Current Game Score**, **Boat Health**, as 3 narrative-based events, have been recorded during the game play (Table 4).

1.2.2. Auditory

Among the controllable incidents, the **Torpedo Approaching** incident creates and alarm sound indicating the presents and distance of an incoming torpedo. This auditory game event can be recorded to identify the moments that the torpedo alarm is on. The sound of explosion can also be considered as another important game event, while it is related to 4 different controllable or normal incidents. The torpedo and mine detonation create an explosion sound. Moreover, the collision of the boat to the invisible barriers would also create an explosion sound. Furthermore, the ship on fire, at the finish line, also creates repetitive explosion sounds in the game play. To distinguish them, 2 game events are recorded while the user is playing the game; **Explosion**, recording mines, torpedo and invisible barriers collision explosions sounds and **Ship on Fire**, recording the repetitive finish line ship's explosions sounds. Despite of the continues sea and wave sounds, which is present in all games, there are 3 other sounds that can be temporally heard. **Score Collection**, **Ramp Jump** and **Water Splash** sounds can be heard when (respectively) a life tube (score) is collected, the boat drives on a ramp to jump and the boat hits the water surface, after a jump. These 3 auditory-related incidents are also recorded in game playtime alongside the other events (Table 4).

1.2.3. Visualisation

Torpedo Approaching can also create a visualisation game event, while the torpedo is in sight of the player's view. This event is recorded during the game run time. The camera shaking and blurring controllable incidents would change the user visualisation. Therefore the **Camera Blurring and Vibration** game event was recorded during the game. On the other hand, in mine avoidance scenarios, the user may decide to go outside the normal boat track and reach the finish line. If this scenario is played in this manner, the user would not encounter any mine, as he/she is trying to avoid the mine field itself. To detect whether the user is experiencing high precision driving within the mine field, the **Number of Surrounding Mines** was calculated in real time and recorded as a game event. Depending on how far the boat is from the finish line, the ship near the finish line flag may be visible by the user. Therefore, the **Distance to Finish Line** and **Ship on Fire Visible** auditory-related game events are also calculated and recorded in game run time (Table 4).

1.2.4. Interactive

In the interactive controllable incidents, if the user is using the joystick with force feedback, he/she will experience haptic interaction with the controller. Therefore, **Vibration** and **Handle Force** are recorded as game events during the gameplay. Moreover, whether the controller works normally or faulty, the user would have different controlling experiences. Therefore, the **Controller Functionality Status** records this game event in playtime. The controller movement (regardless of the type – either mouse or joystick) could also be recorded as an interactive game event. Therefore the **Controller Throttle**, which records the speed of the boat (received from either mouse or joystick), and the **Controller Movement**, which records the right or left commands received from the user, are also recorded as interactive game events (Table 4).

1.2.5. Game Events List

Table 4 shows all recorded game events, which have been calculated and recorded during the game playtime, alongside their format and possible values. This table can be used as a guide, to identify the game events that have been calculated and recorded in real time, alongside the other physiological signals, during the experiment run time.

Table 4 – Target Game Events to be Calculated and Recorded and their Value Guide

Category	Variables	Variable Description
Narrative	Current Game Score	Score Value
	Presented Time	Time in Seconds
	Health of the Boat	Health in %
Auditory	Scoring Collection	0 => No Ding-Ding Sound 1 => Ding-Ding Sound Playing
	Ramp Jump	0 => No Ramp Jumping Sound 1 => Ramp Jumping Sound Playing
	Water Splash	0 => No Water Splash Sound 1 => Water Splash Sound Playing
	Explosion	0 => No Explosion Sound 1 => Explosion Sound Playing
	Torpedo Approaching	0 => No Alarm 1 => Discrete Alarm 2 => Continuous Alarm 3 => White Noise Sound
	Ship on Fire	0 => No Fire or Explosion Sound 1 => Fire Sound 2 => Fire and Explosion Sound Playing
Visualisation	Camera Blurring and Vibrating	0 => Normal Camera 1 => Camera Blurred and Vibrating
	Incoming Torpedo in Sight	0 => No Torpedo in Sight 1 => At least one Torpedo in Sight
	Number of Surrounding Mines	Number of Surrounding Mine
	Ship on Fire in Sight	0 => The Giant Ship out of Sight 1 => The Giant Ship in Sight
	Distance to Finish Line	Distance to Finish Line in % %0 => At the Start Line %100 => At the Finish Line
Interactive	Controller Movement	Position in % Negative => Left Positive => Right
	Controller Throttle	Position in % Negative => Backward Positive => Forward
	Vibration	Strength in %
	Handle Force	Strength in % Negative => Left Positive => Right
	Controller Functionality Status	0 => Working Faulty 1 => Working Normally

2. Real Time Synchronisation

As discussed in Sections 1.1 and 1.2, there are 3 independent recording systems (Shimmer, EMOTIV EPOC and PC to record game events) that are going to be employed to record all physiological signals and game events. Each of these systems employs its own internal timer, to mark each sample with the corresponding time stamp. To be able to relate the recording of these systems, a time synchronisation process has to be conducted. To perform the time synchronisation, 2 distinct stages, covered in Sections 2, need to be conducted.

2.1. Synchronisation Stages

2.1.1. Stage 1 – Start Time Synchronisation

Consider System A and B, recording packages with their own time stamps. Now lets consider that the timing of each of these two systems, is absolutely reliable and has an ignorable non-linear drag or error respect to each other. Now consider the two sets of data shown in Table 5.

Table 5 – Data Time Stamp Synchronisation Example

	Package 0		**Package 1**		**Package 2**		**Package 3**		**Package 4**	
	Data	**Time**	**Data**	**Time**	**Data**	**Time**	**Data**	**Time**	**Data**	**Time**
System A	A_0	TA_0	A_1	TA_1	A_2	TA_2	A_3	TA_3	A_4	TA_4
System B	B_0	TB_0	B_1	TB_1	B_2	TB_2	B_3	TB_3	B_4	TB_4

The first data sets that have arrive from both system A and B are labelled with “0” index. Equation 1 shows the relationship between each time stamp and the *starting time stamp* TX_0. The starting time stamp (TX_0) is the first time stamp that each recording system generates as soon as a recording session is started.

$$TA_n = TA_0 + \left(n \times \frac{1}{fs_A}\right) \quad , \quad fs_A = System\ A\ Sampeling\ Frequency$$

$$TB_n = TB_0 + \left(n \times \frac{1}{fs_B}\right) \quad , \quad fs_B = System\ B\ Sampeling\ Frequency$$

Equation 1 – Independent Recorded Data Timing Formula

As it can be seen in Equation 1, the starting time stamps are the systems’ off-set that could divide the systems’ time instances. Even if the start commands, for both systems, arrive exactly at the same time, the relationship between TA_0 and TB_0 can vary, as each system can have its unique start-up period. Therefore it can be concluded that it is most likely to have $TA_0 \neq TB_0$ rather than $TA_0 = TB_0$. Now consider a particular time in the data recording for both systems, such that $TA_n = TB_n$. Now if $TA_0 = TB_0$ then the corresponding packages from both systems are representing the same time in the recording process. But if $TA_0 > TB_0$, then $A_n$was recorded after B_n, and vice versa. This issue concludes that the first synchronisation stage should be carried out to calculate the starting time stamp for each system. To do so, two approaches can be selected.

1. **One of the Systems as the Reference System:** In this approach one of the systems is chosen to act as the reference timer. Then the TX_0 of the other systems can be calculated respect to the receive time of the packages according to the chosen reference system. It has to be remembered that in this approach the fastest system, which always starts sooner than the others, has to be selected as the reference system. As an illustration, imagine that System A is chosen to act as the reference system. Then at time TA_m the first data arrives from system B. In that situation the TB_0 should be equal to TA_m, while TA_0 is set to zero. Now the relationship, shown in Equation 2, present the links between TA_n and TB_n as their starting time has been synchronised according to System A.

$$TA_n = \left(n \times \frac{1}{fs_A}\right) \quad , \quad fs_A = System\ A\ Sampeling\ Frequency$$

$$TB_n = TA_m + \left(n \times \frac{1}{fs_B}\right) \quad , \quad fs_B = System\ B\ Sampeling\ Frequency$$

Equation 2 – Dependent, Device Related, Recorded Data Timing Formula

2. **Master System as the Reference System:** In this approach the start time of both systems will be synched according to a master system. To do so, a master system, which controls both System A and B, would be considered as the reference timer for both systems. Imagine TM_x represents the master time stamps. Now consider that the first data package arrives from System A at TM_{A0}, and then from System B at TM_{B0}. Therefore Equation 3 would govern the relationship between the master and system A and B time stamps.

$$TA_n = TM_{A0} + \left(n \times \frac{1}{fs_A}\right) \quad , \quad fs_A = System\ A\ Sampeling\ Frequency$$

$$TB_n = TM_{B0} + \left(n \times \frac{1}{fs_B}\right) \quad , \quad fs_B = System\ B\ Sampeling\ Frequency$$

Equation 3 – Dependent, Master System Related, Recorded Data Timing Formula

2.1.2. Stage 2 – Sampling Time Synchronisation

As it was explained in Section 2.1.1, the starting times could be synchronised, according to 2 distinct approaches, to imply common meanings at time stamps of all systems. However, respect to various sampling frequencies, the systems' time stamps can be separated as well. There are 4 types of time separation scenarios (discussed below). Consider the same systems introduced at Table 5.

1. **Type 1:** One of the systems is considered as the reference system (approach 1 in Section 2.1.1 – Equation 2 should be applied), and both systems are operating

with equal sampling frequencies. Therefore Equation 4 is going to govern both systems' timing.

$$TA_n = TB_k$$
$$\left(n \times \frac{1}{fs_A}\right) = TA_m + \left(k \times \frac{1}{fs_B}\right)$$
$$fs_A = fs_B \quad => \quad \frac{1}{fs_A} = \frac{1}{fs_B} = Ts$$
$$(n-k)\times Ts = TA_m$$
$$k = n - m \quad as \quad TA_m = m\times Ts$$

Equation 4 – Type 1 Time Synchronisation Process – n and k are Particular Indices in Package A and B

2. **Type 2:** One of the systems is considered as the reference system (approach 1 in Section 2.1.1 – Equation 2 should be applied), while each system is operating with different sampling frequencies. Therefore if $fs_B = l\times fs_A$, then "l" should be an integer number to enable the overall system to have a k^th element in B for each n^th element in A. Equation 5 shows the mathematical calculations.

$$TA_n = TB_k$$
$$\left(n \times \frac{1}{fs_A}\right) = TA_m + \left(k \times \frac{1}{fs_B}\right)$$
$$fs_A \neq fs_B \quad => \quad Ts_A = \frac{1}{fs_A} \quad , \quad Ts_B = \frac{1}{fs_B}$$
$$nTs_A - kTs_B = TA_m$$
$$if \quad fs_B = l\times fs_A \quad then \quad Ts_B = \frac{Ts_A}{l}$$
$$n - \frac{k}{l} = \frac{TA_m}{Ts_A} \quad , \quad TA_m = m\times Ts_A \quad => \quad n - \frac{k}{l} = m$$
$$k = (n-m)\times l \quad => l\ has\ to\ be\ an\ integer\ to\ generate\ integer\ k$$

Equation 5 – Type 2 Time Synchronisation Process – n and k are Particular Indices in Package A and B

3. **Type 3:** The Master system is considered as the reference system (approach 2 in Section 2.1.1 – Equation 3 should be applied), and both systems are operating with equal sampling frequencies. Equation 6 shows the mathematical calculations to find the k^th element in B, which shares the same time stamp with n^th package in A. Equation 6 shows that unlike Type 1 and 2, the k^th element is going contain an error element. If ΔT is equal to zero, it means that system B is started exactly at a time instance of system A ($TM_{B0} = TA_m$). In this case the systems would be governed according to Equation 5 rather than Equation 6. But if ΔT is non-zero (which is mostly likely to happen in majority of cases – No TA_m exist for TM_{B0}), then it means that the first sample of system B has happened in between 2 samples of system A ($TB_0 \neq TA_m$ or $TB_0 \neq TA_{m+1}$).

As the systems are going to be synched together then to identify the k^th^ sample in B, which shares the same time stamp in A, TB_0 should be pushed either back to TA_m or forward to TA_{m+1}. This can be decided according to the amount of error (options 1 or 2 in Equation 6). In this case there would be always a small amount of error in the system, which is equal to $\pm\frac{Ts}{2}$.

$$TA_n = TB_k$$

$$TM_{A0} + \left(n \times \frac{1}{fs_A}\right) = TM_{B0} + \left(k \times \frac{1}{fs_B}\right)$$

$$fs_A = fs_B \quad => \quad \frac{1}{fs_A} = \frac{1}{fs_B} = Ts$$

$$(n-k) \times Ts = TM_{B0} - TM_{A0}$$

$$k = n + \frac{TM_{A0} - TM_{B0}}{Ts}$$

$$TM_{A0} - TM_{B0} = z \times Ts + \Delta T\ , means{:}\ z\ sampels\ plus\ a\ \Delta T\ , while \quad 0 < \Delta T < Ts$$

$$then{:}\ k = n + \frac{TM_{B0} - TM_{A0}}{Ts} = n + z + \frac{\Delta T}{Ts}$$

$$\frac{\Delta T}{Ts} = ERROR \ => \ 0 < ERROR < 1$$

$$k = n + z + ERROR$$

1) $\quad 0 < \Delta T < \frac{Ts}{2} \ => \ k = n + z$

2) $\quad \frac{Ts}{2} \le \Delta T < Ts \ => \ k = n + z + 1$

Equation 6 – Type 3 Time Synchronisation Process – n and k are Particular Indices in Package A and B

4. **Type 4:** The Master system is considered as the reference system (approach 2 in Section 2.1.1 – Equation 3 should be applied), while each system is operating with different sampling frequencies. This case is similar to Type 3, and the only difference is that the systems are not working with the same speed. As it can be obtained from Equation 7, the only difference is that the k^th^ element in B, which shares the same time stamp in n^th^ package in A, is "l" times larger than the Equation 6, and this is due to this fact that system B is now "l" times faster than system A. But as it can be seen the error element is still in the equation and causes some inaccuracy in synchronisation process, in terms of $\pm\frac{Ts_A}{2}$ second.

$$TA_n = TB_k$$

$$TM_{A0} + \left(n \times \frac{1}{fs_A}\right) = TM_{B0} + \left(k \times \frac{1}{fs_B}\right)$$

$$fs_A \neq fs_B \quad => \quad Ts_A = \frac{1}{fs_A} \quad , \quad Ts_B = \frac{1}{fs_B}$$

$$n \times Ts_A - k \times Ts_B = TM_{B0} - TM_{A0}$$

$$if \quad fs_B = l \times fs_A \quad then \quad Ts_B = \frac{Ts_A}{l}$$

$$k = l \times \left(n + \frac{TM_{A0} - TM_{B0}}{Ts_A} \right)$$

$$TM_{B0} - TM_{A0} = z \times Ts_A + \Delta T \text{ }, means: z \text{ } sampels \text{ } plus \text{ } a \text{ } \Delta T \text{ }, while \quad 0 < \Delta T < Ts_A$$

$$then: k = l \times \left(n + \frac{TM_{A0} - TM_{B0}}{Ts_A} \right) = l \times \left(n + z + \frac{\Delta T}{Ts_A} \right)$$

$$\frac{\Delta T}{Ts_A} = ERROR \;\; => \;\; 0 < ERROR < 1$$

$$k = l \times (n + z + ERROR) => l \text{ } has \text{ } to \text{ } be \text{ } an \text{ } integer \text{ } to \text{ } generate \text{ } integer \text{ } k$$

1) $0 < \Delta T < \frac{Ts_A}{2} \;\; => \;\; k = l \times (n + z)$

2) $\frac{Ts_A}{2} \leq \Delta T < Ts_A \;\; => \;\; k = l \times (n + z + 1)$

Equation 7 – Type 4 Time Synchronisation Process – n and k are Particular Indices in Package A and B

2.2. Experiment Devices Synchronisation

In this experiment, the EMOTIV EPOC and Shimmer+ can be controlled by their system development kits; meaning that the start and stop command can be initiated at any instances. However, the system responses can vary according to different hardware and software (SDK) processes. In this case, there is almost no way to guarantee that the systems start working at the same time instance. Therefore, the main system (the mother software – Section 3.2) would acts as the master system to assign the starting time of the each system accordingly, as soon as each system is started (approach 2 in Section 2.1.1 – Equation 3 should be applied).

On the other hand, the EMOTIV EPOC comes with a single non-adjustable sampling frequency, which is 128Hz. The Shimmer+, on the other hand, comes with 8 adjustable sampling frequencies. The game events can be recorded with any required frequencies (Section 3.5). Table 6 shows the availability of specific sampling frequencies for each system. The game event recording speed was set to be equal to the EMOTIV EPOC sampling rate (128Hz). As Shimmer could not operate with a 128Hz sampling frequency option, the smallest available option, which is "l" (an integer value) times larger than 128Hz sampling rate, is 512Hz. Therefore the Type 4 synchronisation process (Equation 7) has been employed in the time stamp alignment.

Table 6 – Shimmer, EMOTIV EPOC and Game Event Data Recording Sampling Frequency choices

128Hz x 4 = 512Hz

Possible Sampeling Frequencies	1Hz	10.2Hz	51.2Hz	102.4Hz	128Hz	204.8Hz	250Hz	512Hz	1024Hz
Shimmer	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Epoc Emotiv	No	No	No	No	Yes	No	No	No	No
Game Events	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

3. Recording Software Structure

As it was explained earlier, the recording program is responsible of recording and synchronising the required signals from 3 systems; EMOTIV EPOC, Shimmer and the game event recorder (see section 1.1 and 1.2.5). One of the system requirements is to allow the experiment supervisor to control and monitor the performance of the systems. While the participant is playing the game, the screen is completely occupied with the game display, and any additional information, which is irrelevant to the game itself and would assist the supervisor, would distract the participant, and disturb the affective stimulation process. There are some system parameters that need to be monitored during the experiment run time, which would help the supervisor to ensure that the performance of the recording system is appropriate. In case of any error or problem (system stops saving the data into the file due to an internal error, one of the devices disconnects, etc.), the supervisor should be able to stop the game from the monitoring system, and fix the issues before restarting the game again. This will ensure that the database is recorded completely without any data missing or recording corruption. This Section describes the designing stages of this complex system.

In the experiment, there are 3 major operations that have to be performed simultaneously; the game software, the recording process and the monitoring system. Each of these systems has to be able to be run separately, without causing any delay or disturbance on the other. As it was explained earlier, the monitoring system needs to be run separately from the game environment, and any interaction with the monitoring system should not create any disturbance on the participants. Therefore, a tablet-based application has been designed to monitor the recording process, through a wireless (wifi) network. Figure 4 and Table 7 present the recording system's black box diagram and symbol guide. As it can be seen from the diagram, there are 5 sub-systems that are running separately. Each sub-system is described briefly in Sections 3.2 to 3.9.

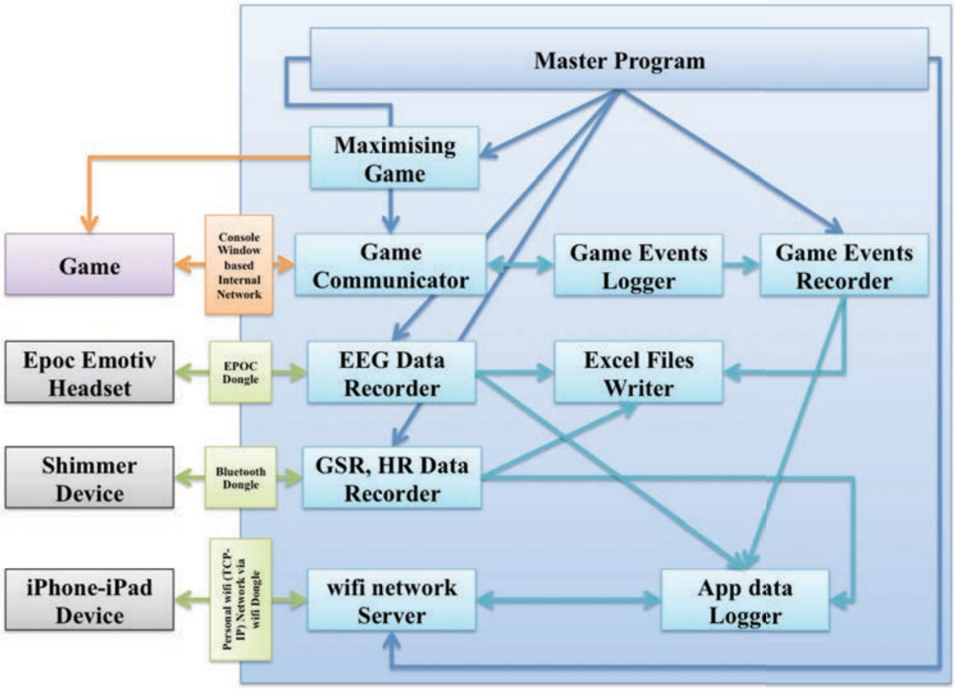


Figure 4 – Physiological Signal Recording Experiment, System Black-Box Diagram

Table 7 – Physiological Signal Recording Experiment, Black-Box Diagram Symbol Guide

Symbol	Description
	Internal communication between programs, through TCP-IP socket network
	External communication between PC and devices, through Wifi or Bluetooth devices
	Recording program's master threads or connections
	Multi/parallel threads within the main recording program
	External devices (Epoc Emotiv, Shimmer and iPad)
	The game program

3.1. .NET (C#) Multi Threading Algorithm

Multithreading and threaded programing is usually used in multitasks systems. In this programing technique the independent tasks, called *threads*, which does not require the completion of another to be completed, could be run simultaneously. One of the most important features of this technique is the ability of using single resource for multiple threads in a process at the same time (as long as there is no access interference, such as updating a field while another is trying to access it). To achieve multithreading in a process, the process has to be carefully segmented (threaded) to perform the tasks, which can be handled independently and simultaneously. Then an instance of the required threads can be run, controlled and terminated in parallel with a number of other threads, within a process. The multithreading feature can have several advantages, which some of them have been addressed below[10] (Agafonov, 2013).

1. **Responsiveness:** In processes that the system has to be responsive to multiple input/output interfaces, this ability enables the systems to stay active and handle all interfaces, without creating any delays, queues or freezing situations, on the others. Moreover the system can handle intercommunications, calculations and data manipulations, while other threads manage interfaces simultaneously without causing any delays.
2. **Faster Execution:** Segmenting the tasks and operating them simultaneously, speed up the overall process considerably, as the tasks do not cause any queuing delays to be completed.
3. **Lower Resources (Memory) Consumption:** In this approach, multiple threads can share the required resources and minimise the memory usage. During memory and resources sharing, it should be remembered that no conflicts between the threads could happen. As an illustration, if two threads try to modify or update some parameters at the same time, a data hazard can occur. To prevent conflict hazards the "**lock**" statement[11] can be used to queue any simultaneous access to specific parameters, according to their time of access (first attempt is allowed to have the first access).

The most important drawback of multithreading, is the synchronisation process. As in multitask processes, which use multithreading algorithms, tasks have to be executed at specific timings, which can be controlled by either the main (master) process, or another thread. Moreover the data accessing timings and intercommunications also needs to be synched to enable the pipeline (the overall structure of the system) to perform appropriately. Furthermore, the termination of some threads has to be monitored or in some cases forced, to prevent any system freezing incidents. To achieve these demands, the .NET "Thread" class[12] has been used to execute and terminate the threaded tasks (see Figure 4) within the recording process.

[10] Joseph Albahari, Threading in C# Online Book: http://www.albahari.com/threading/
[11] https://msdn.microsoft.com/en-us/library/c5kehkcz.aspx
[12] https://msdn.microsoft.com/en-us/library/system.threading(v=vs.110).aspx

3.2. Master Program

This sub-system carries the main structure of the recording process. This section would control all threads and parallel programs that would be simultaneously running alongside of the main system. Also this sub-system would ensure that all connections have been established. Moreover, in case of receiving the termination command, this section makes sure that all remaining data in RAM is copied into the corresponding Excel file, before performing any overall termination.

3.3. Game Application

The game is an independent application that is run at the beginning of the experiment. This application is responsible to run the affective VR experience, as described in Chapter 3. The game would start the recording application (.exe file) within itself, as a new controllable process, each time a game session is started (not in the introduction scenes, emotion assessment pages and the training sessions). The .NET "Process" class[13] has been employed to start, control and terminate the recording program, remotely, from the game software. The game monitors the status of the recording program. In case of any unexpected termination, the program would stop the game and return to the diagnoses page (Section 3.10). Moreover, when the game is going to be finished (after passing the finish line), a termination command will be sent to the recording program. This will allow the recording program to stop the recording process, copy all the remaining data into the storage files, and terminate itself.

As soon as the recording application is executed by the game software, the master program minimises the console windows and maximises the game (make sure the game is always at the top – the Maximising Game sub-system – Section 3.4). The console window is accessible by both programs (the game and the recording program – The Game Communicator sub-system). This would provide a platform for both systems to communicate with each other through string-based commands. As during the run time the game acts as the master system, while the recording program as the slave, all commands would come from the game software. A commands-table has been created in each system to allow both systems to send and interpret string-based commands appropriately.

3.4. Maximising the Game

This part of the system would be run as a parallel thread at the beginning of the program. The section is responsible to make sure that the game interface always stays as the top window and other programs are considered as background applications. This sub-system ensures that the gameplay is not interrupted due to any minimisation action and the screen is always occupied with the game interface.

[13] https://msdn.microsoft.com/en-us/library/System.Diagnostics.Process(v=vs.110).aspx

3.5. Game Events Recorder

As it was explained earlier, there are 19 game evens that need to be recorded (see section 1.2.5). The game itself is rendered by the game engine during the run time, in 25 to 50 frame per-second. Within each frame the game events can be updated. However, the frame-rendering rate can vary according to the system specification (GPU and CPU), and the game environment. As an illustration, the frame rate is higher in "Free Environment Exploring" scenarios, when compared to the "Torpedo Avoidance" scenario (according to high processing requirement). As the game events have to be calculated and recording at each frame, then the sampling frequency for game events would vary between 25 to 50 Hz. To solve this issue a thread has been created to sample the game event data pool (Game Events Logger) within the system. The data pool sampling frequency has been fixed at 128Hz, while the data pool itself would be updated according to the game frame rendering frequency. This would fix the variability and speed issues in the recording process.

To calculate the game events, the event recorder program within the game (which has access to the corresponding objects, which carry the required events within the game) would calculate the game events status in each rendered frame and transfer them to the recorder program through the recording program console window (Section 3.3).

3.6. EEG Data Recorder

This sub-system is responsible of establishing a connection with EMOTIV EPOC headset and start receiving the data packages. The EMOTIV EPOC provides a .NET based (C#) system development kit (SDK) to provide the required functions to establish connection with the headset and provide the communication protocols to receive data packages. This sub-system will be executed as a parallel thread and would establish connection to the headset. After that it would receive a package of 128 samples, in every second. The samples within the packages are time stamped with the EPOC sampling time. After establishing connection, the system time[14] is used to save the system recording start time (Section 2). After that all samples' time stamps are recalculated according to the system and device start time to create the correct, aligned and synched time stamp, for each sample. Equation 8 shows the synchronisation process that is carried out by the EEG recording thread. The packages with recalculated time stamps are stored in EEG temporarily data-pool to be stored into Excel files accordingly. EMOTIV EPOC employed an internal algorithm to provide signal quality measures. These quality measures have been employed in the feature matrix construction process, to discard the low quality portions.

$$TD_i => The\ original\ i^{th}\ sample\ time\ received\ from\ \boldsymbol{EPOC\ Headset}$$
$$TS_0 => The\ recording\ start\ time, based\ on\ the\ \boldsymbol{system}\ time$$
$$TD_0 => The\ time\ of\ the\ first\ received\ sample\ from\ \boldsymbol{EPOC\ Headset}$$
$$TC_i => The\ corrected\ i^{th}\ sample\ time$$

[14] https://msdn.microsoft.com/en-us/library/System.DateTime(v=vs.110).aspx

$$(TD_i - TD_0) = k \times T_S => k^{th}\ sample\ after\ recording\ start\ time$$
$$TC_i = TS_0 + (TD_i - TD_0) = TS_0 + (k \times T_S) - (see\ \text{Equation 7})$$

Equation 8 – EEG Time Calculation Formula

3.7. GSR, PPG and HR Data Recorder

This sub-system is responsible for establishing a connection with the Shimmer+ device and start receiving the data packages. The Shimmer+ provides a .NET based (C#) system development kit (SDK) to provide the required functions to establish connection with the device and provide the communication protocols to receive data packages. As soon as the recording is started, the device starts recording with the specified sampling frequency (512Hz). The device marks the recorded data with the device internal timer and transfers the time stamped package through a Bluetooth connection. Each package would be received and stored into a temporarily data-pool to be stored in the Excel files accordingly. The timing of each package would be recalculated according to Equation 8 algorithm.

The Shimmer+ provides "-1" for the heart rate during the periods, which either the PPG to heart rate conversion calibration process are carried out, or the PPG signal quality is low. However no quality measures have been provided by Shimmer+ or implemented by the authors to investigate the quality of the GSR signals.

3.8. Excel File Writer

This sub-system has access to "Game Data Recorder", "EEG Data Recorder", "GSR, PPG and HR Data Recorder" and the temporarily data-pools. This thread would be run as a parallel thread from the "Recorder" threads, and copy the data into the corresponding Excel files and clear the threads temporarily data-pools. In case of any Excel file corruption or access denial, the thread creates another Excel file and copies the future data into the new file. This makes sure that no data is lost due to any file corruption, and the data can be used later by concatenating the Excel files to each other.

3.9. App Data Logger

This sub-system contains the most recent information about the recording process. This information would be sent to the monitoring app (by the "wifi Network Server" thread), through a personal wifi network. The purpose of the monitoring app is to enable the supervisor, to inspect the performance of the each session individually in real time to be able to restart the session or adjust some parameters, in case of any recording disturbance or sudden system failure.

3.9.1. Wifi Network Server

This sub-system is run as a parallel thread at the start of the recording program. As it was stated earlier, the most important requirement of this system is to enable the supervisor to monitor the system performance, without causing any distortion in both game experiencing and recording processes. Therefore, this thread would work independently, from the recording threads, and in case of any failure or absence of any table connection, it does not cause any interrupt or disturbance in other threads. This thread would open a TCP-IP protocol based socket through the active wifi network (using Microsoft Windows ad-hoc wifi network[15]) as a host, and waits for incoming client to get connected. In case of a connection request, the client would be considered as the monitoring app and the content of the "App Data Logger" would be transferred regularly to the app, through text-based messages. To achieve these demands, the .NET "Socket" class[16] has been used to manage the connection and data transfer process. The "Wifi Network Server" thread can receive only the restart command from the app client. In case of receiving this command, the thread informs the game and the recording master program, to shut down the recording process and terminate the game and move to a diagnoses page (see section 3.10).

3.9.2. Tablet (iPad) Application

To design and implement the monitoring app on an ios-based tablet (iPad), the XCode software[17] has been employed. Moreover to be able to install and run the designed app on an iPad, a developer account was purchased through the Apple Company website[18]. The application establishes a TCP-IP protocol based connection through the main wifi with the computer. Then it receives system performance parameters, from the recording program ("wifi Network Server" thread), to update some visual information on the app screen, to enable the supervisor to assess the performance of the recording system.

Figure 5 shows an example of the iPad application screen in the run time. As it can be seen the subject heart rate is 67 beats per-minute. Also both Shimmer and EPOC headset are connected properly and have fairly strong signals. The data streaming is in progress for both devices, however the channels signal quality for some of the EPOC electrodes are not acceptably high (Green colour).

The supervisor is able to use the app tutorial, designed within the application, to review the PC-iPad connection process. During the Process the IPV4 address (Internet Protocol Version 4 – TCP-IP addressing Protocol) of the network should be identified and added to the app's interface. Figure 6 shows the IPV4 identification process, and inserting the retrieved address into the iPad application to establish the wireless connection.

The storyboard of the app behaviour is created within the XCode software to manage the performance of the application accordingly. Figure 7 shows the designed and implemented storyboard within the XCode software.

---

[15] http://windows.microsoft.com/en-gb/windows/set-computer-to-computer-adhoc-network#1TC=windows-7
[16] https://msdn.microsoft.com/en-us/library/system.net.sockets(v=vs.110).aspx
[17] https://developer.apple.com/xcode/
[18] https://developer.apple.com/

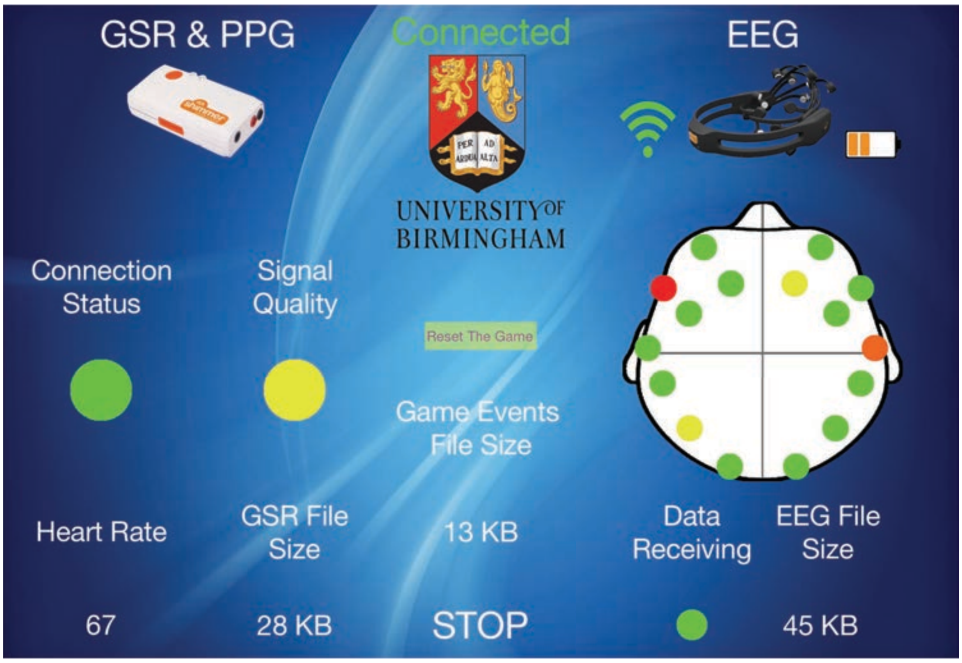


Figure 5 – iPad App Monitoring Screen Example

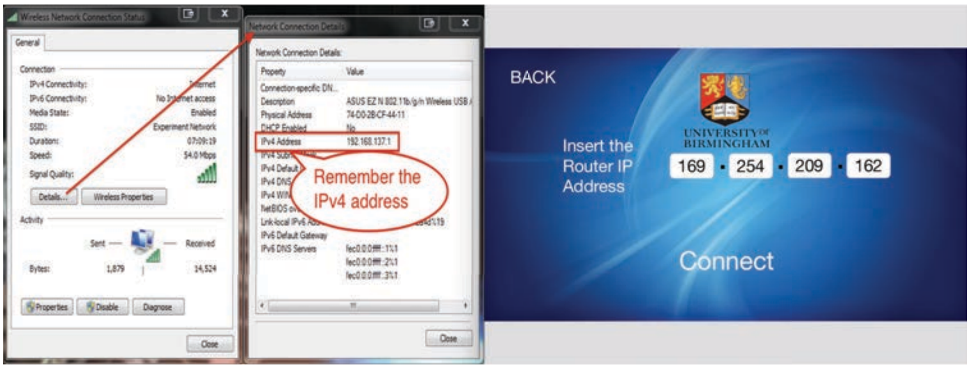


Figure 6 – Identification and Implementation of IPV4 Address

Figure 7 – iPad Application Storyboard

3.10. Diagnoses Page

This system is specifically designed to report the causes behind a session crash. In case of any unexpected errors or sudden termination in the system, unsuccessful device connection attempt, any noticeable low signal quality in the EEG channels, or a restart command received from the monitoring app, the game would be terminated and this page would be presented. Then the source(s) of the failure would be presented on the screen for the supervisor to handle the situation and restart the session after solving the problem. Figure 8 and Figure 9 show 2 examples of the possible errors and the diagnoses reports.

Figure 8 – Diagnoses Page Example 1 – Failed to connect to Shimmer and EMOTIV EPOC Headset

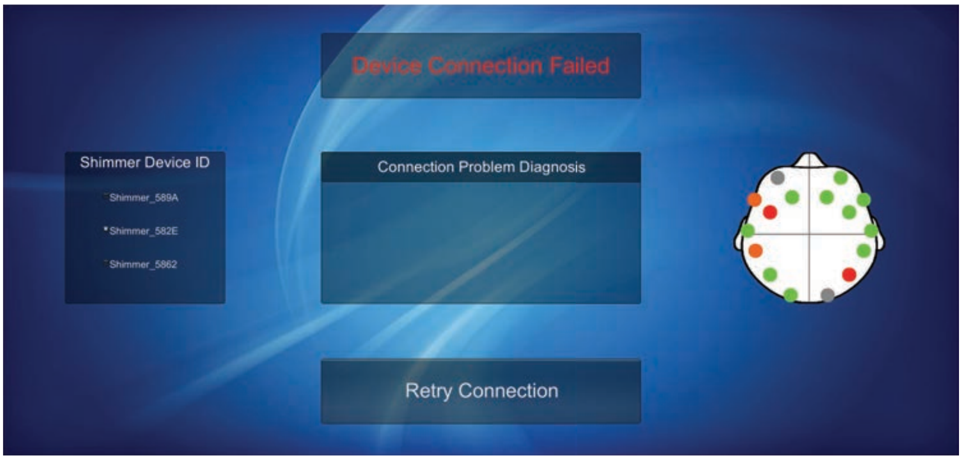


Figure 9 – Diagnoses Page Example 2 – Low signal quality on some EMOTIV EPOC Channels

References

Agafonov, E., 2013. *Multithreading in C# 5.0 Cookbook*. Packt Publishing.

K. Nakajima, T.T.H.M., 1996. Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique. *Medical Engineering & Physics*, 18(5), pp.365–72.

Appendix I

Adaptive Virtual Environments:
A Psychophysiological Feedback HCI System Concept

Experiment Purpose:

This experiment is designed to assess the emotional impact of different games. You will be presented with 10 games, each of which last between 2-5 minutes and has its own characteristics and objectives. You are asked to fill in a questioner after each game to assess your emotional experience during the game. Meanwhile, some physiological signals of your body will be recorded (EEG from skull, Heart-Rate by PPG and GSR (Galvanic Skin Response) on your figures). At the end of experiment a £10 Amazon voucher would be presented.

Experiment:

The background scenario for all of the games will stay the same. In the beginning of the experiment a resting state (means that you are relaxed and not highly active) of your body will be recorded. Then you will pass a training session to be familiarised with the games. The objectives, skills and environment will be introduced during the training session. The game score will be calculated at the end of each game. You are supposed to gain the highest possible score for each game.

Although each game can have its own tasks and objectives, all games can be **finished** by passing through the **"Finish Line"** flag. In games that have time limitation, you **<u>have to</u>** pass through the "Start Line" flag to start the timer, otherwise the game score will be calculated as zero. Also if you lose all of the **boat health**, the game will be finished with zero score. Moreover, If you do not manage to finish the game in the limited time provide, the game will be finished with zero score. If you spend more than 5 minutes in the games with **no time limitation** (means you do not pass the finish line by then), the game will be automatically ended, and your score will be calculated.

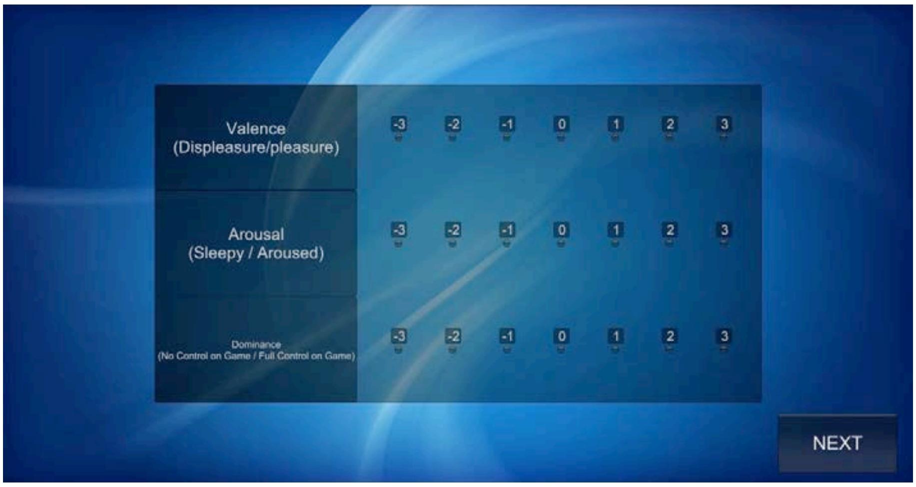
Before each game, an introduction page will introduce the overall and important game's tasks, objectives and controller. Please **consider** each of them carefully; otherwise you will **miss** the objective of the game. Also during some games some or all parts of the game may behave incorrectly:

1. **Faulty Timer**: The timer can be faulty and randomly increase the time.
2. **Faulty Controller**: The controller also can stop responding, or even behave wrongly (ex. Reverse the controller axis).
3. **Invisible Barrier**: There may be invisible barriers in the game with stop signs on them. In case of getting too close to them, they will turn visible and reduce the boat health in case of any collision.

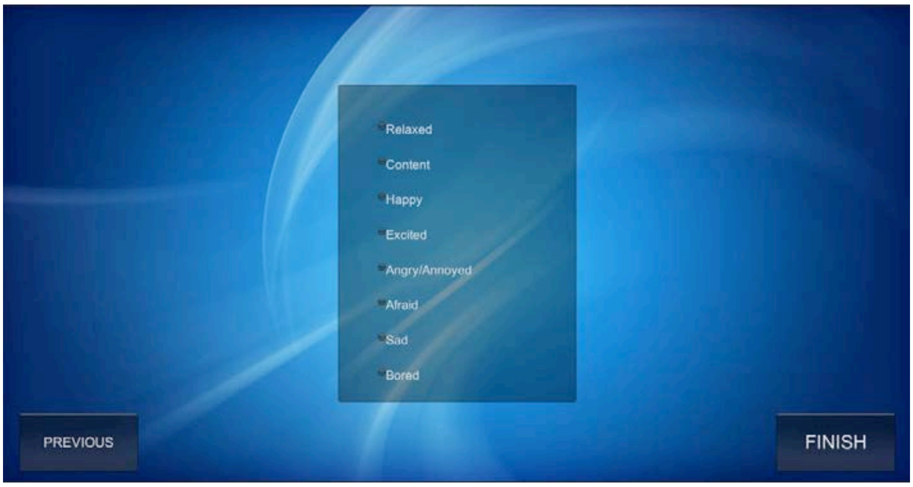
Questioner:

At the end of each game you are asked to rate your emotional experience during the game. The ratings are based on 3 scales:

1. **<u>Valence</u>**: How pleasurable this game experience was. Higher positive value means more pleasure (you like it) and higher negative value means displeasure (you dislike it).
2. **<u>Arousal</u>**: How arousing this game experience was. Higher positive value means more aroused (e.g. excited, alert, stressful, etc.) and higher negative value means negatively aroused (e.g. relaxed, sleepy, tired, bored, etc.).
3. **<u>Dominance</u>**: How much control you have on the game. Higher positive value means more control on the game and higher negative value means the game is out of your control.



In addition, we ask you to choose the most appropriate emotion label, from a list, that fits the game experience.



Please try to assess your emotional experience **<u>during</u>** the game, and try **<u>not</u>** to answer the questioner based on the outcome of the game.

Experiment Duration:

The experiment will last at most 100 minutes. After each questioner, you can rest as long as you prefer and start the next game as soon as seems appropriate. Although, we will appreciate if you finish the experiment completely, you are allowed to leave the experiment after each questioner at any stage.

Physiological Signal Recording:

All Physiological signals are measured by non-invasive devices. This means that all the recording devices are passive and **<u>do not</u>** create any direct contact with the nerve system. Although there will be a break in the middle of the experiment and the devices will be removed from your body, **<u>please</u>** inform the instructor at any point that you felt discomfort with any of the devices.

Appendix J

	Subject				Session Information											Affective Rating					
Feature Number =>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Index Number	Subject Number	Gender	Age	Hand Preference	Session Number	Game Code Digit 1	Game Code Digit 2	Game Code Digit 3	Game Code Digit 4	Game Code Digit 6	Game Code Digit 7	Game Code Digit 8	Game Code Digit 9	Game Start Time	Session Duration	Valance	Arousal	Dominance	Emotion Label	Theta Power Summation	Slow Alpha Power Summation
0	1 - 30				Resting State	Resting State	Resting State	Resting State	Resting State	Resting State	Resting State	Resting State	Resting State	Float Value in Seconds	Float Value in Seconds	Integer Number From -3 to 3	Integer Number From -3 to 3	Integer Number From -3 to 3		Float Value	Float Value
1		Male	12 - 18	Right	Others => 1 - 10	Free Environment Exploring	Time Limitation	Normal Timer	No Invisible Barrier	Mouse	Normal Controller	Shaking and Blurring The Camera	Colour Screen						Relaxed		
2		Female	18 - 24	Left		Mines Avoidance	No Time Limitation	Faulty Timer	Invisible Barrier	Joystick Without Force Feedback	Faulty Controller	No Camera Shaking or Blurring	Black & White Screen						Content		
3			24 - 30			Torpedo Avoidance				Joystick With Force Feedback			Inverse Black & White Screen						Happy		
4			30 - 40			Shooting Flying Ball													Excited		
5						Maze													Angry / Annoyed		
6																			Afraid		
7																			Sad		
8																			Bored		

Single Channel AF3

Feature Number =>	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Index Number	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					

<table>
<tr><td></td><td colspan="6"></td><td colspan="15">Single Channe</td></tr>
<tr><td>Feature Number =></td><td>43</td><td>44</td><td>45</td><td>46</td><td>47</td><td>48</td><td>49</td><td>50</td><td>51</td><td>52</td><td>53</td><td>54</td><td>55</td><td>56</td><td>57</td><td>58</td><td>59</td><td>60</td><td>61</td><td>62</td><td>63</td></tr>
<tr><td>Index Number</td><td>Beta Power RMS Radio db / Alpha Power RMS Radio db</td><td>EEGw (Power Summation)</td><td>EEGw (Power Ratio)</td><td>EEGw (Power RMS)</td><td>abs(abs(EEGw (Power RMS Ratio db)))</td><td>Average Quality</td><td>Theta Power Summation</td><td>Slow Alpha Power Summation</td><td>Alpha Power Summation</td><td>Beta Power Summation</td><td>Gamma Power Summation</td><td>Theta Power Ratio</td><td>Slow Alpha Power Ratio</td><td>Alpha Power Ratio</td><td>Beta Power Ratio</td><td>Gamma Power Ratio</td><td>Theta Power RMS</td><td>Slow Alpha Power RMS</td><td>Alpha Power RMS</td><td>Beta Power RMS</td><td>Gamma Power RMS</td></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td>No Signal</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td><td>Very Bad</td></tr>
<tr><td>2</td><td>Poor</td></tr>
<tr><td>3</td><td>Fair</td></tr>
<tr><td>4</td><td>Good</td></tr>
<tr><td>5</td><td rowspan="4">Float Value Between 0 and 4</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

l AF4

Feature Number =>	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
Index Number	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1														Very Bad							
2														Poor							
3														Fair							
4														Good							
5														Float Value Between 0 and 4							
6																					
7																					
8																					

<table>
<tr><th></th><th colspan="21">Single Channel F3</th></tr>
<tr><th>Feature Number =></th><th>85</th><th>86</th><th>87</th><th>88</th><th>89</th><th>90</th><th>91</th><th>92</th><th>93</th><th>94</th><th>95</th><th>96</th><th>97</th><th>98</th><th>99</th><th>100</th><th>101</th><th>102</th><th>103</th><th>104</th><th>105</th></tr>
<tr><th>Index Number</th><th>Alpha Power Ratio</th><th>Beta Power Ratio</th><th>Gamma Power Ratio</th><th>Theta Power RMS</th><th>Slow Alpha Power RMS</th><th>Alpha Power RMS</th><th>Beta Power RMS</th><th>Gamma Power RMS</th><th>Theta Power RMS Ratio db</th><th>Slow Alpha Power RMS Ratio db</th><th>Alpha Power RMS Ratio db</th><th>Beta Power RMS Ratio db</th><th>Gamma Power RMS Ratio db</th><th>Beta Power Summation / Alpha Power Summation</th><th>Beta Power Ratio / Alpha Power Ratio</th><th>Beta Power RMS / Alpha Power RMS</th><th>Beta Power RMS Radio db / Alpha Power RMS Radio db</th><th>EEGw (Power Summation)</th><th>EEGw (Power Ratio)</th><th>EEGw (Power RMS)</th><th>abs(abs(EEGw (Power RMS Ratio db)))</th></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td></tr>
<tr><td>2</td></tr>
<tr><td>3</td></tr>
<tr><td>4</td></tr>
<tr><td>5</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

		Single Channel F4																			
Feature Number =>	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126
Index Number	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db
0	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1	Very Bad																				
2	Poor																				
3	Fair																				
4	Good																				
5	Float Value Between 0 and 4																				
6																					
7																					
8																					

<table>
<tr><td>Feature Number =></td><td>127</td><td>128</td><td>129</td><td>130</td><td>131</td><td>132</td><td>133</td><td>134</td><td>135</td><td>136</td><td>137</td><td>138</td><td>139</td><td>140</td><td>141</td><td>142</td><td>143</td><td>144</td><td>145</td><td>146</td><td>147</td></tr>
<tr><td>Index Number</td><td>Beta Power Summation / Alpha Power Summation</td><td>Beta Power Ratio / Alpha Power Ratio</td><td>Beta Power RMS / Alpha Power RMS</td><td>Beta Power RMS Radio db / Alpha Power RMS Radio db</td><td>EEGw (Power Summation)</td><td>EEGw (Power Ratio)</td><td>EEGw (Power RMS)</td><td>abs(abs(EEGw (Power RMS Ratio db)))</td><td>Average Quality</td><td>Theta Power Summation</td><td>Slow Alpha Power Summation</td><td>Alpha Power Summation</td><td>Beta Power Summation</td><td>Gamma Power Summation</td><td>Theta Power Ratio</td><td>Slow Alpha Power Ratio</td><td>Alpha Power Ratio</td><td>Beta Power Ratio</td><td>Gamma Power Ratio</td><td>Theta Power RMS</td><td>Slow Alpha Power RMS</td></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td>No Signal</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td><td>Very Bad</td></tr>
<tr><td>2</td><td>Poor</td></tr>
<tr><td>3</td><td>Fair</td></tr>
<tr><td>4</td><td>Good</td></tr>
<tr><td>5</td><td rowspan="4">Float Value Between 0 and 4</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

Feature Number =>	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
	Single Channel F7																				
Index Number	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value
1																	Very Bad				
2																	Poor				
3																	Fair				
4																	Good				
5																	Float Value Between 0 and 4				
6																					
7																					
8																					

<table>
<tr><td></td><td colspan="21">Single Channel F8</td></tr>
<tr><td>Feature Number =></td><td>169</td><td>170</td><td>171</td><td>172</td><td>173</td><td>174</td><td>175</td><td>176</td><td>177</td><td>178</td><td>179</td><td>180</td><td>181</td><td>182</td><td>183</td><td>184</td><td>185</td><td>186</td><td>187</td><td>188</td><td>189</td></tr>
<tr><td>Index Number</td><td>Gamma Power Summation</td><td>Theta Power Ratio</td><td>Slow Alpha Power Ratio</td><td>Alpha Power Ratio</td><td>Beta Power Ratio</td><td>Gamma Power Ratio</td><td>Theta Power RMS</td><td>Slow Alpha Power RMS</td><td>Alpha Power RMS</td><td>Beta Power RMS</td><td>Gamma Power RMS</td><td>Theta Power RMS Ratio db</td><td>Slow Alpha Power RMS Ratio db</td><td>Alpha Power RMS Ratio db</td><td>Beta Power RMS Ratio db</td><td>Gamma Power RMS Ratio db</td><td>Beta Power Summation / Alpha Power Summation</td><td>Beta Power Ratio / Alpha Power Ratio</td><td>Beta Power RMS / Alpha Power RMS</td><td>Beta Power RMS Radio db / Alpha Power RMS Radio db</td><td>EEGw (Power Summation)</td></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td></tr>
<tr><td>2</td></tr>
<tr><td>3</td></tr>
<tr><td>4</td></tr>
<tr><td>5</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

					Single Channel FC5																
Feature Number =>	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
Index Number	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db
0	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1				Very Bad																	
2				Poor																	
3				Fair																	
4				Good																	
5				Float Value Between 0 and 4																	
6																					
7																					
8																					

Feature Number =>	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231
Index Number	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1												Very Bad									
2												Poor									
3												Fair									
4												Good									
5												Float Value Between 0 and 4									
6																					
7																					
8																					

	Single Channel FC6																				
Feature Number =>	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252
Index Number	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value
1																				Very Bad	
2																				Poor	
3																				Fair	
4																				Good	
5																				Float Value Between 0 and 4	
6																					
7																					
8																					

Single Channel T7

Feature Number =>	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273
Index Number	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					

																					Singl
Feature Number =>	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294
Index Number	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1							Very Bad														
2							Poor														
3							Fair														
4							Good														
5							Float Value Between 0 and 4														
6																					
7																					
8																					

	e Channel T8																				
Feature Number =>	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315
Index Number	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio
0															No Signal						
1															Very Bad						
2															Poor						
3															Fair						
4	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Good	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
5															Float Value Between 0 and 4						
6																					
7																					
8																					

Single Channel P7

Feature Number =>	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336
Index Number	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					

			Single Channel P8																		
Feature Number =>	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357
Index Number	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db
0	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1		Very Bad																			
2		Poor																			
3		Fair																			
4		Good																			
5		Float Value Between 0 and 4																			
6																					
7																					
8																					

	EEG Features																				
Feature Number =>	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378
Index Number	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1										Very Bad											
2										Poor											
3										Fair											
4										Good											
5										Float Value Between 0 and 4											
6																					
7																					
8																					

<table>
<tr><td></td><td colspan="18">Single Channel O1</td><td colspan="3"></td></tr>
<tr><td>Feature Number =></td><td>379</td><td>380</td><td>381</td><td>382</td><td>383</td><td>384</td><td>385</td><td>386</td><td>387</td><td>388</td><td>389</td><td>390</td><td>391</td><td>392</td><td>393</td><td>394</td><td>395</td><td>396</td><td>397</td><td>398</td><td>399</td></tr>
<tr><td>Index Number</td><td>Slow Alpha Power RMS</td><td>Alpha Power RMS</td><td>Beta Power RMS</td><td>Gamma Power RMS</td><td>Theta Power RMS Ratio db</td><td>Slow Alpha Power RMS Ratio db</td><td>Alpha Power RMS Ratio db</td><td>Beta Power RMS Ratio db</td><td>Gamma Power RMS Ratio db</td><td>Beta Power Summation / Alpha Power Summation</td><td>Beta Power Ratio / Alpha Power Ratio</td><td>Beta Power RMS / Alpha Power RMS</td><td>Beta Power RMS Radio db / Alpha Power RMS Radio db</td><td>EEGw (Power Summation)</td><td>EEGw (Power Ratio)</td><td>EEGw (Power RMS)</td><td>abs(abs(EEGw (Power RMS Ratio db)))</td><td>Average Quality</td><td>Theta Power Summation</td><td>Slow Alpha Power Summation</td><td>Alpha Power Summation</td></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td>No Signal</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td><td>Very Bad</td></tr>
<tr><td>2</td><td>Poor</td></tr>
<tr><td>3</td><td>Fair</td></tr>
<tr><td>4</td><td>Good</td></tr>
<tr><td>5</td><td rowspan="4">Float Value Between 0 and 4</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

	Single Channel O2																				
Feature Number =>	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420
Index Number	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					

<table>
<tr><td></td><td colspan="5"></td><td colspan="16">Paired Electrodes AF3 - A</td></tr>
<tr><td>Feature Number =></td><td>421</td><td>422</td><td>423</td><td>424</td><td>425</td><td>426</td><td>427</td><td>428</td><td>429</td><td>430</td><td>431</td><td>432</td><td>433</td><td>434</td><td>435</td><td>436</td><td>437</td><td>438</td><td>439</td><td>440</td><td>441</td></tr>
<tr><td>Index Number</td><td>EEGw (Power Summation)</td><td>EEGw (Power Ratio)</td><td>EEGw (Power RMS)</td><td>abs(abs(EEGw (Power RMS Ratio db)))</td><td>Average Quality</td><td>Theta Power Summation</td><td>Slow Alpha Power Summation</td><td>Alpha Power Summation</td><td>Beta Power Summation</td><td>Gamma Power Summation</td><td>Theta Power Ratio</td><td>Slow Alpha Power Ratio</td><td>Alpha Power Ratio</td><td>Beta Power Ratio</td><td>Gamma Power Ratio</td><td>Theta Power RMS</td><td>Slow Alpha Power RMS</td><td>Alpha Power RMS</td><td>Beta Power RMS</td><td>Gamma Power RMS</td><td>Theta Power RMS Ratio db</td></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td>No Signal</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td><td>Very Bad</td></tr>
<tr><td>2</td><td>Poor</td></tr>
<tr><td>3</td><td>Fair</td></tr>
<tr><td>4</td><td>Good</td></tr>
<tr><td>5</td><td rowspan="4">Float Value Between 0 and 4</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

F4

Feature Number =>	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462
Index Number	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1													Very Bad								
2													Poor								
3													Fair								
4													Good								
5													Float Value Between 0 and 4								
6																					
7																					
8																					

	Paired Electrodes F3 - F4																				
Feature Number =>	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483
Index Number	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal
1																					Very Bad
2																					Poor
3																					Fair
4																					Good
5																					Float Value Between 0 and 4
6																					
7																					
8																					

Paired Electrodes F7 - F8

Feature Number =>	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504
Index Number	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					

Pa

Feature Number =>	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
Index Number	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1								Very Bad													
2								Poor													
3								Fair													
4								Good													
5								Float Value Between 0 and 4													
6																					
7																					
8																					

	ired Electrodes FC5 - FC6																				
Feature Number =>	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546
Index Number	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))	Gamma Power RMS Ratio db	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value
1																Very Bad					
2																Poor					
3																Fair					
4																Good					
5																Float Value Between 0 and 4					
6																					
7																					
8																					

Paired Electrodes T7 - T8

Feature Number =>	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567
Index Number	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio	Theta Power RMS	Slow Alpha Power RMS	Alpha Power RMS	Beta Power RMS	Gamma Power RMS	Theta Power RMS Ratio db	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					

<table>
<tr><td></td><td colspan="3"></td><td colspan="18">Paired Electrodes P7 - P8</td></tr>
<tr><td>Feature Number =></td><td>568</td><td>569</td><td>570</td><td>571</td><td>572</td><td>573</td><td>574</td><td>575</td><td>576</td><td>577</td><td>578</td><td>579</td><td>580</td><td>581</td><td>582</td><td>583</td><td>584</td><td>585</td><td>586</td><td>587</td><td>588</td></tr>
<tr><td>Index Number</td><td>EEGw (Power RMS)</td><td>abs(abs(EEGw (Power RMS Ratio db)))</td><td>Average Quality</td><td>Theta Power Summation</td><td>Slow Alpha Power Summation</td><td>Alpha Power Summation</td><td>Beta Power Summation</td><td>Gamma Power Summation</td><td>Theta Power Ratio</td><td>Slow Alpha Power Ratio</td><td>Alpha Power Ratio</td><td>Beta Power Ratio</td><td>Gamma Power Ratio</td><td>Theta Power RMS</td><td>Slow Alpha Power RMS</td><td>Alpha Power RMS</td><td>Beta Power RMS</td><td>Gamma Power RMS</td><td>Theta Power RMS Ratio db</td><td>Slow Alpha Power RMS Ratio db</td><td>Alpha Power RMS Ratio db</td></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td>No Signal</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td><td>Very Bad</td></tr>
<tr><td>2</td><td>Poor</td></tr>
<tr><td>3</td><td>Fair</td></tr>
<tr><td>4</td><td>Good</td></tr>
<tr><td>5</td><td rowspan="4">Float Value Between 0 and 4</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

Feature Number =>	589	590	591	592	593	594	595	596	597	598
Index Number	Beta Power RMS Ratio db	Gamma Power RMS Ratio db	Beta Power Summation / Alpha Power Summation	Beta Power Ratio / Alpha Power Ratio	Beta Power RMS / Alpha Power RMS	Beta Power RMS Radio db / Alpha Power RMS Radio db	EEGw (Power Summation)	EEGw (Power Ratio)	EEGw (Power RMS)	abs(abs(EEGw (Power RMS Ratio db)))
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1										
2										
3										
4										
5										
6										
7										
8										

Feature Number =>	599	600	601	602	603	604	605	606	607	608	609
Index Number	Average Quality	Theta Power Summation	Slow Alpha Power Summation	Alpha Power Summation	Beta Power Summation	Gamma Power Summation	Theta Power Ratio	Slow Alpha Power Ratio	Alpha Power Ratio	Beta Power Ratio	Gamma Power Ratio
0	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1	Very Bad										
2	Poor										
3	Fair										
4	Good										
5	Float Value Between 0 and 4										
6											
7											
8											

<table>
<tr><td></td><td colspan="18">Paired Electrodes O1 - O2</td><td colspan="3">Paired Electroc</td></tr>
<tr><td>Feature Number =></td><td>610</td><td>611</td><td>612</td><td>613</td><td>614</td><td>615</td><td>616</td><td>617</td><td>618</td><td>619</td><td>620</td><td>621</td><td>622</td><td>623</td><td>624</td><td>625</td><td>626</td><td>627</td><td>628</td><td>629</td><td>630</td></tr>
<tr><td>Index Number</td><td>Theta Power RMS</td><td>Slow Alpha Power RMS</td><td>Alpha Power RMS</td><td>Beta Power RMS</td><td>Gamma Power RMS</td><td>Theta Power RMS Ratio db</td><td>Slow Alpha Power RMS Ratio db</td><td>Alpha Power RMS Ratio db</td><td>Beta Power RMS Ratio db</td><td>Gamma Power RMS Ratio db</td><td>Beta Power Summation / Alpha Power Summation</td><td>Beta Power Ratio / Alpha Power Ratio</td><td>Beta Power RMS / Alpha Power RMS</td><td>Beta Power RMS Radio db / Alpha Power RMS Radio db</td><td>EEGw (Power Summation)</td><td>EEGw (Power Ratio)</td><td>EEGw (Power RMS)</td><td>abs(abs(EEGw (Power RMS Ratio db)))</td><td>Average Quality</td><td>Slow Alpha Power Summation</td><td>Alpha Power Summation</td></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td>No Signal</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td><td>Very Bad</td></tr>
<tr><td>2</td><td>Poor</td></tr>
<tr><td>3</td><td>Fair</td></tr>
<tr><td>4</td><td>Good</td></tr>
<tr><td>5</td><td rowspan="4">Float Value Between 0 and 4</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

	les Power Spectral Analysis of AF3 & AF4 => (Pleft - Pright) / (Pleft + Pright)							Paired Electrodes Power Spectral Analysis of F3 & F4 => (Pleft - Pright) / (Pleft + Pright)									Paired Electrodes Power Spectral Ana Pright) / (Pleft +				
Feature Number =>	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651
Index Number	Slow Alpha Power Ratio	Alpha Power Ratio	Slow Alpha Power RMS	Alpha Power RMS	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Average Quality	Slow Alpha Power Summation	Alpha Power Summation	Slow Alpha Power Ratio	Alpha Power Ratio	Slow Alpha Power RMS	Alpha Power RMS	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Average Quality	Slow Alpha Power Summation	Alpha Power Summation	Slow Alpha Power Ratio	Alpha Power Ratio	Slow Alpha Power RMS
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value
1							Very Bad									Very Bad					
2							Poor									Poor					
3							Fair									Fair					
4							Good									Good					
5							Float Value Between 0 and 4									Float Value Between 0 and 4					
6																					
7																					
8																					

	lysis of F7 & F8 => (Pleft - Pright)				Paired Electrodes Power Spectral Analysis of FC5 & FC6 => (Pleft - Pright) / (Pleft + Pright)									Paired Electrodes Power Spectral Analysis of T7 & T8 => (P							
Feature Number =>	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672
Index Number	Alpha Power RMS	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Average Quality	Slow Alpha Power Summation	Alpha Power Summation	Slow Alpha Power Ratio	Alpha Power Ratio	Slow Alpha Power RMS	Alpha Power RMS	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Average Quality	Slow Alpha Power Summation	Alpha Power Summation	Slow Alpha Power Ratio	Alpha Power Ratio	Slow Alpha Power RMS	Alpha Power RMS	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db
0	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1				Very Bad									Very Bad								
2				Poor									Poor								
3				Fair									Fair								
4				Good									Good								
5				Float Value Between 0 and 4									Float Value Between 0 and 4								
6																					
7																					
8																					

	left -	Paired Electrodes Power Spectral Analysis of P7 & P8 => (Pleft - Pright) / (Pleft + Pright)									Paired Electrodes Power Spectral Analysis of O1 & O2 => (Pleft - Pright) / (Pleft + Pright)									Left Frontal E	
Feature Number =>	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693
Index Number	Average Quality	Slow Alpha Power Summation	Alpha Power Summation	Slow Alpha Power Ratio	Alpha Power Ratio	Slow Alpha Power RMS	Alpha Power RMS	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Average Quality	Slow Alpha Power Summation	Alpha Power Summation	Slow Alpha Power Ratio	Alpha Power Ratio	Slow Alpha Power RMS	Alpha Power RMS	Slow Alpha Power RMS Ratio db	Alpha Power RMS Ratio db	Average Quality	Power Summation	Power Ratio
0	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value
1	Very Bad									Very Bad									Very Bad		
2	Poor									Poor									Poor		
3	Fair									Fair									Fair		
4	Good									Good									Good		
5	Float Value Between 0 and 4									Float Value Between 0 and 4									Float Value Between 0 and 4		
6																					
7																					
8																					

	EGw => AF3, F3, F7, FC5			Right Frontal EEGw => AF4, F4, F8, FC6					Left Parietal EEGw => P7, O1					Right Parietal EEGw => P8, O2					Frontal EEGw => AF3 F7, F8, FC5,		
Feature Number =>	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714
Index Number	Power RMS	abs(Power RMS Ratio db)	Average Quality	Power Summation	Power Ratio	Power RMS	abs(Power RMS Ratio db)	Average Quality	Power Summation	Power Ratio	Power RMS	abs(Power RMS Ratio db)	Average Quality	Power Summation	Power Ratio	Power RMS	abs(Power RMS Ratio db)	Average Quality	Power Summation	Power Ratio	Power RMS
0	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value	Float Value	No Signal	Float Value	Float Value	Float Value
1			Very Bad					Very Bad					Very Bad					Very Bad			
2			Poor					Poor					Poor					Poor			
3			Fair					Fair					Fair					Fair			
4			Good					Good					Good					Good			
5			Float Value Between 0 and 4					Float Value Between 0 and 4					Float Value Between 0 and 4					Float Value Between 0 and 4			
6																					
7																					
8																					

<table>
<tr><td></td><td colspan="2">, AF4, F3, F4, FC6</td><td colspan="5">Parietal EEGw => P7, P8, O1, O2</td><td colspan="5">Overall EEGw</td><td colspan="9"></td></tr>
<tr><td>Feature Number =></td><td>715</td><td>716</td><td>717</td><td>718</td><td>719</td><td>720</td><td>721</td><td>722</td><td>723</td><td>724</td><td>725</td><td>726</td><td>727</td><td>728</td><td>729</td><td>730</td><td>731</td><td>732</td><td>733</td><td>734</td><td>735</td></tr>
<tr><td>Index Number</td><td>abs(Power RMS Ratio db)</td><td>Average Quality</td><td>Power Summation</td><td>Power Ratio</td><td>Power RMS</td><td>abs(Power RMS Ratio db)</td><td>Average Quality</td><td>Power Summation</td><td>Power Ratio</td><td>Power RMS</td><td>abs(Power RMS Ratio db)</td><td>Average Quality</td><td>Mean</td><td>Minimum</td><td>Maximum</td><td>Standard Deviation</td><td>Mean of The Peaks</td><td>Mean of the First Derivative</td><td>Mean of the Positive Values in First Derivative</td><td>Mean of the Negative Values in First Derivative</td><td>Mean of The First Derivative Peaks</td></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td>No Signal</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td>No Signal</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td>No Signal</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td><td>Very Bad</td><td>Very Bad</td><td>Very Bad</td></tr>
<tr><td>2</td><td>Poor</td><td>Poor</td><td>Poor</td></tr>
<tr><td>3</td><td>Fair</td><td>Fair</td><td>Fair</td></tr>
<tr><td>4</td><td>Good</td><td>Good</td><td>Good</td></tr>
<tr><td>5</td><td rowspan="4">Float Value Between 0 and 4</td><td rowspan="4">Float Value Between 0 and 4</td><td rowspan="4">Float Value Between 0 and 4</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

	Heart Rate													O2 SatuRatio						
Feature Number =>	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755
Index Number	Medium Frequency Power Summation	High Frequency Power Summation	Medium Frequency Power Ratio	High Frequency Power Ratio	Medium Frequency RMS	High Frequency RMS	Medium Frequency RMS Ratio db	High Frequency Power Ratio db	Medium Power Summation / High Power Summation	Medium Power Ratio / High Power Ratio	Medium RMS / High RMS	Medium RMS Ratio db / High RMS Ratio db	Fluctuation Frequency	Mean	Mean	Minimum	Maximum	Standard Deviation	Mean of The Peaks	Mean of the First Derivative
0	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Integer Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value	Float Value
1																				
2																				
3																				
4																				
5																				
6																				
7																				
8																				

<table>
<tr><td></td><td colspan="8">GSR</td><td>Time</td></tr>
<tr><td>Feature Number =></td><td>756</td><td>757</td><td>758</td><td>759</td><td>760</td><td>761</td><td>762</td><td>763</td><td>764</td></tr>
<tr><td>Index Number</td><td>Mean of the Positive Values in First Derivative</td><td>Mean of the Negative Values in First Derivative</td><td>Mean of The First Derivative Peaks</td><td>Low Frequency Power Summation</td><td>Low Frequency Power Squred Sumation</td><td>Low Frequency RMS</td><td>Low Frequency RMS Ratio db</td><td>Fluctuation Frequency</td><td>Window Central Time</td></tr>
<tr><td>0</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9">Float Value</td><td rowspan="9"></td><td rowspan="9">Float Value</td></tr>
<tr><td>1</td></tr>
<tr><td>2</td></tr>
<tr><td>3</td></tr>
<tr><td>4</td></tr>
<tr><td>5</td></tr>
<tr><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
</table>

Appendix K

Iterative Feature Selection Technique: A PCA-Based Feature Selection

1. Features Selection

As discussed in Chapter 5, in total, 743 features have been extracted from the physiological raw signals. To be able to perform emotion classification, the dimension of the features matrix has to be reduced to a subspace. This subspace has fewer dimensions (called *Most Optimal Features* throughout the appendix), while it can adequately capture the essence of the data (Murphy, 2012). To do so the following steps have been conducted.

1.1. Principal Component Analysis (PCA)

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. Also consider the emotion vector $E \in \mathbb{R}^{n}$, while n is the number of observations. Consider that f_{ij} is the j^{th} feature of the i^{th} observation, and e_i presents the observation's corresponding emotional label (either in dimensional or categorical model). A matrix $Z \in \mathbb{R}^{n \times l}$ could be constructed, such that Z is the rotated version of F, while n and l are the number of observations and variables, respectively. Therefore a conversion matrix $W \in \mathbb{R}^{d \times l}$ needs to be constructed, such that $Z^T = W \times F^T$. The Principal Component Analysis (PCA) finds the most optimal W matrix, which can convert the features to a set of linearly uncorrelated variables, called principal components (Murphy, 2012). Each principal component (z_{ik}) can be constructed, by a linear mixture of all corresponding features (Equation 1 – z_{ik} is the k^{th} principal component of the i^{th} observation). Moreover, the PCA algorithm guarantees that matrix W is an *Orthonormal* matrix ($WW^T = I$). This means that the rows of matrix W are linearly independent. This concludes that all w_{kj} within a row (corresponding weights of z_{ik}) could be linearly compared, while they cannot be compared to other weights, in other rows of matrix W, as they are linearly independent.

On the other hand, if l is equal to d, meaning that the number of constructed principal components is equal to the number of features, then both $F \text{ and } Z \in \mathbb{R}^{n \times d}$. In that case the W matrix will be a square matrix, such that $W \in \mathbb{R}^{d \times d}$. Now if the rows of W are sorted in a descending order, according to the eigenvalues of the covariance matrix of the features matrix F[1]; the columns of the principal component matrix Z would be in order of decreasing components' variance. This guarantees that the first principal component would have the maximum variance compared to the others; the second principal component would have a larger variance compared to the

[1] Meaning that the row, which has the larger eigenvalue, would be the first row of the W and so on.

others, except the first component, and so on (Murphy, 2012). The PCA analysis was implemented using MATLAB software (version R2015b), which provides the W matrix sorted according to the eigenvalues, in a descending format.

$$\begin{bmatrix} z_1 \\ \vdots \\ z_d \end{bmatrix}_{d\times 1} = \begin{bmatrix} w_{11} & \cdots & w_{1d} \\ \vdots & \ddots & \vdots \\ w_{d1} & \cdots & w_{dd} \end{bmatrix}_{d\times d} \times \begin{bmatrix} f_1 \\ \vdots \\ f_d \end{bmatrix}_{d\times 1}$$

$$Z = \begin{bmatrix} z_1 = \sum_{i=1}^{d} w_{1i} \times f_i \\ \vdots \\ z_d = \sum_{i=1}^{d} w_{di} \times f_i \end{bmatrix}$$

Equation 1 – Features to Principals Components Conversion Using PCA Rotation Matrix "*W*"

1.2. Sparse PCA

The corresponding weights for each principal component is a value between -1 and 1, and could indicate the contribution of each feature, within the construction of each principal component. As an illustration, if the weights of features f_n and f_m, for construction of principal component z_k, are -0.8 and 0.1 respectively, it could be concluded that the contribution of f_n, in construction of z_k, is considerably more than f_m. A "PCA Weight Threshold" function was defined to set all w_{ld}, which their absolute values is smaller than a threshold value (β – Equation 2), to zero and the rest to one (similar to the Sparse PCA algorithm applied by (Cadima & Jolliffe, 1995)). As shown in Equation 2, matrix S is a binary matrix, identifying the features, which mostly contribute to each principal component by one, and the rest by zero (considering the contribution threshold β). Matrix S is called the **Sparse PCA** matrix. Therefore, the first row of S ($[s_{11} \quad \cdots \quad s_{1d}]$) presents the features' contribution pattern to the most important principal component, while the second row ($[s_{21} \quad \cdots \quad s_{2d}]$) presents the features' contribution pattern to the second most important PC, and so on.

$$S = \begin{bmatrix} s_{11} & \cdots & s_{1d} \\ \vdots & \ddots & \vdots \\ s_{d1} & \cdots & s_{dd} \end{bmatrix} = Threshold \left\{ \begin{bmatrix} w_{11} & \cdots & w_{1d} \\ \vdots & \ddots & \vdots \\ w_{d1} & \cdots & w_{dd} \end{bmatrix}, \beta \right\}$$

$$S_{(i\times j)} = Threshold\{A_{(i\times j)}, \beta\} = \begin{cases} 0, & |a_{ij}| < \beta \\ 1, & |a_{ij}| \geq \beta \end{cases}$$

Equation 2 – PCA Threshold Function

1.3. Correlation of Features

As discussed in Chapter 5, several measurements for a single physiological signal have been extracted. As an illustration, minimum, maximum and average heart rate are extracted from each window, while these measurements are highly correlated. As another example, measuring the frequency bandwidth power using all four equations, presented in Chapter 5, creates highly correlated features, as they are all different measures of a single variable. This issue concludes the fact that the features matrix carries a number of highly correlated features. As a numerical examples, the average GSR values and the mean of the GSR peaks in windows are significantly ($P<0.001$) correlated, with 0.97 as the correlation coefficient. Or the Alpha rhythm power in channel O1, measured with the summation formula is significantly ($P<0.001$) correlated with the same power measured in RMS, with 0.9 as the correlation coefficient.

These high correlation within the features matrix, means that a high number of features can be considered as redundant, as the other correlated counterparts could be predicted without direct measurement. Therefore, selecting a fewer number of uncorrelated features can reduce the dimensionality of the features matrix considerably.

1.4. Iterative Feature Selection

To design a technique, which can reduce the dimensionality of the dataset and select the most relevant features, that could provide sufficient information to classify the physiological affective space, the following requirements have been defined:

1. **Contribution to Principal Components:** The selected features must have the highest contributions to the most important principal components. This means that the features, which contribute more to the first principal component, have a higher priority to be selected in the feature selection process, compare to the features, which contribute more to the second component.
2. **Minimum Correlation:** The selected features should have minimum correlation with each other. This is due to the fact that selecting two highly correlated features creates redundancy in the feature space, as the other correlated counterparts, can be predicted without direct measurement.
3. **Adjustable Number of Features:** The algorithm has to be able to select the d most relevant features (adjustable, from 1 to D).

To be able to satisfy these requirements, an iterative feature selection technique has been design. This algorithm is constructed of two separate steps; (1) feature sorting according to their contribution to the principal components and (2) rejecting the correlated features.

1.4.1. Feature Sorting

Equation 3 presents the pseudo codes of the feature sorting process. First all features, which their corresponding values in the first row of the sparse PCA matrix S are one, are identified and added to an empty matrix P. Then Matrix P is sorted, in a descending order, according to first row of matrix W, such that, the features with higher absolute weights (w_{1j}) appear first, compared to those, which have smaller absolute weights. After that matrix P is appended to matrix I, which carries the final selected features. In the second step, all features, which their corresponding values in the second row of the sparse PCA matrix S are one, are identified and added to the empty matrix P, if they have not been selected in the previous iteration (not in matrix I). Then the sorted matrix P is added to matrix I. This process is repeated until all rows of the sparse PCA matrix S are analysed.

for i=1,…,d
for j=1,…,d
if ((s_{ij} = 1) AND ($f_j \notin I$))
Add f_j to P
Sort P Respect to abs(W_i)*, in a Descending Order*
Append P to I
Clear P

Equation 3 – Pseudo Codes For Features Sorting According to their Contribution to The PCs

The sorting process, according to the PCA weights' absolute values, is conducted separately at each row of matrix W. This is due to the fact, that the rows of matrix W are linearly independent (Section 1.1), and the sorting have to be conducted independently, within each row of the matrix W. At the end of this process matrix I, which carries all features in order of their contributions to the principal components, would be created. This means that the first feature in matrix I is the most important feature, which has the highest contribute to the first principal component. The second feature in matrix I, on the other hand, is less important than the first feature, while more important than the rest of the features, and so on.

1.4.2. Correlated Features Elimination

To eliminate the highly correlated features, another iterative algorithm has been designed and implemented in this study. Equation 4 presents the designed pseudo codes for "highly correlated features" elimination process. As it can be seen in Equation 4, the correlation of all features is assessed in their importance order. This means that, in case there is a high significant correlation between the i^{th} and j^{th} features[2] (larger than or equal a threshold, α), the j^{th} element is eliminated in case j is larger than i. This is due to the fact that, in the sorted feature vector I, the i^{th} feature has a higher contribution to the physiological affective space, than the j^{th} feature, if j is larger than i. Therefore, if there is a high correlation between i^{th} and j^{th} features, the feature with less contribution in the physiological affective space needs to be eliminated. The correlation significance level is evaluated using the *Bonferroni Correction* algorithm (Hommel, 1983). Hence, the correlation of the first most important feature is compared to all others, and in case of a high and significant correlation, the less important feature is eliminated. This comparison is continued, until all features are compared with the others. The final "I" vector contains all minimally correlated features (considering the correlation threshold α), which are sorted in terms of their contributions to the physiological affective space.

[2] The i^{th} and j^{th} feature within vector I.

for all $f_i \in I$, $f_j \in I$

if((abs(corr $(f_i , f_j)) \geq \alpha$ *) AND (Significant Correlation) AND* $(j > i)$ *)*

Remove f_j *from* I

Equation 4 – Pseudo Codes For Eliminating Highly Correlated Features – "*abs*" Means the Absolut Value – "$corr(f_i , f_j)$" Means the Correlation Coefficient of the i^{th} and j^{th} Features – Significant Correlation Means that the Correlation P-Value is Smaller than the Critical Threshold (using Bonferroni algorithm)

1.4.3. Final Feature Selection

Finally the first d features from vector I can be selected, as they are the most important features (maximum contribution to the most important principal components) and minimally correlated, and could provide sufficient information to classify the physiological affective space.

1.5. Iterative Feature Selection Parameters

As it was explained in Sections 1.2, 1.3 and 1.4, there are three parameters in this algorithm, that can be adjusted, in order to select a number of most relevant features, that could provide sufficient information to classify the physiological affective space; (1) PCA Weight Rejection Threshold ($\boldsymbol{\beta}$), (2) Correlation Coefficient Rejection Threshold ($\boldsymbol{\alpha}$) and (3) Number of Required Features ($\boldsymbol{d}$). We arbitrary selected eight values (0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.15, 0.2), for the PCA weight rejection threshold (β). Also, we selected 19 arbitrary correlation coefficient rejection thresholds (α – 0.05 to 1 with step size of 0.05). Moreover, 30 arbitrary values have been used as the number of required features (d – 1 to 30). The combinations can create 4560 different settings ($8\times19\times30 = 4560$) for the iterative variable selection process. The optimal selection of these parameters is investigated in Section 2.1.

2. Hyper-Parameters Tuning

By considering all four classifiers (SVM, DA, Classification Tree and KNN – as presented in Chapter 5), there are around 19.5 million different classifiers[3], which could be trained according to various windowing techniques, feature selection and classification settings. As it is almost impossible to consider all 19.5 million settings, we performed a selective grid search (Bergstra & Bengio, 2012) (as discussed in Chapter 5). In this selective grid search, the feature selection parameters (α as the correlation coefficient threshold and β as the PCA weight threshold) have been assessed through a 100,000-element grid search, employing only KNN classifier. This was due to the fact that, in the physiological database, the KNN classifiers always outperforms other techniques, with higher accuracy and shorter training time (refer to Section 3.2). Then other classifiers have been evaluated, using another 200,000-element grid search, by fixing the feature selection parameters (α and β), identified

[3] $56 \text{ (Window type and length)} \times 4{,}560\ (\alpha, \beta \text{ and D}) = 255{,}360 \text{ (Windowing and Feature Selection Settings)}$
$24 \text{ (SVM)} + 2 \text{ (DA)} + 20 \text{ (Classification Tree)} + 30\ (KNN) = 76\ (Classification\ Sttings)$
$255{,}360 \text{ (Windowing and Feature Selection Settings)} \times 76\ (Classification\ Settings) = 19{,}407{,}360\ (Classifiers)$

from the best performing classifiers, in the former 100,000-element grid search (refer to Section 2.1). As a result, in this study, we covered only (around) 2% of all possible hyper-parameters' variations, to evaluate the performance of the different settings. According to (Bergstra & Bengio, 2012), this small subset of the larger settings space could be sufficient in the hyper-parameters tuning process, as majority of the hyper-parameters variations do no matter much, as only those, which result in high accuracy, matter. To assess the performance of different classifications settings, the accuracies of the classifiers have been estimated through a 10-Fold (random folding) Cross Validation technique (Murphy, 2012).

2.1. Feature Selection Settings Evaluation

As discussed in Section 1.5, the iterative feature selection algorithm can be tuned, using three parameters; α as the correlation coefficient rejection threshold, β as the PCA weight rejection threshold and d as the number of features, employed to perform the classification process. As it was mentioned in Section 2, the α and β parameters are tuned respect to the KNN classifier. Figure 1 presents the performance of the KNN classifier, respect to different values of PCA weight rejection threshold (β). As it can be obtained by the graph, the accuracy of the classifiers is not affected significantly by changing this value (β). The maximum accuracies for all PCA rejection thresholds are around 92% (overall maximum KNN classification accuracy, employing four features – Figure 4). Therefore, the middle value (β=0.1) has been selected for the PCA weight rejection threshold. Figure 2 presents the accuracies of the KNN classifier, respect to different correlation coefficient rejection thresholds. As it can be obtained by the graph, the maximum accuracy of the classifiers increases at α=0.2; while it does not change for larger correlation coefficient rejection thresholds. As it can be seen the maximum accuracy for all α values, larger than 0.2, is around 92% (overall maximum KNN classification accuracy employing **four** features – Figure 4). This is due to the fact that for large α values, the best classifiers employ highly correlated features. These highly correlated features could be almost the mirror of the four most optimal features, which result in 92% classification accuracy. To resolve this issue, we repeated the analysis, using only the classifiers, which employ less than or equal to 4 features (Figure 3). As it can be obtain by Figure 3, the most optimal interval for the correlation coefficient rejection threshold (α) is between 0.2 and 0.55. To ensure that only minimally correlated features are selected in the feature selection algorithm, the smallest possible α (0.2) has been selected as the correlation coefficient threshold.

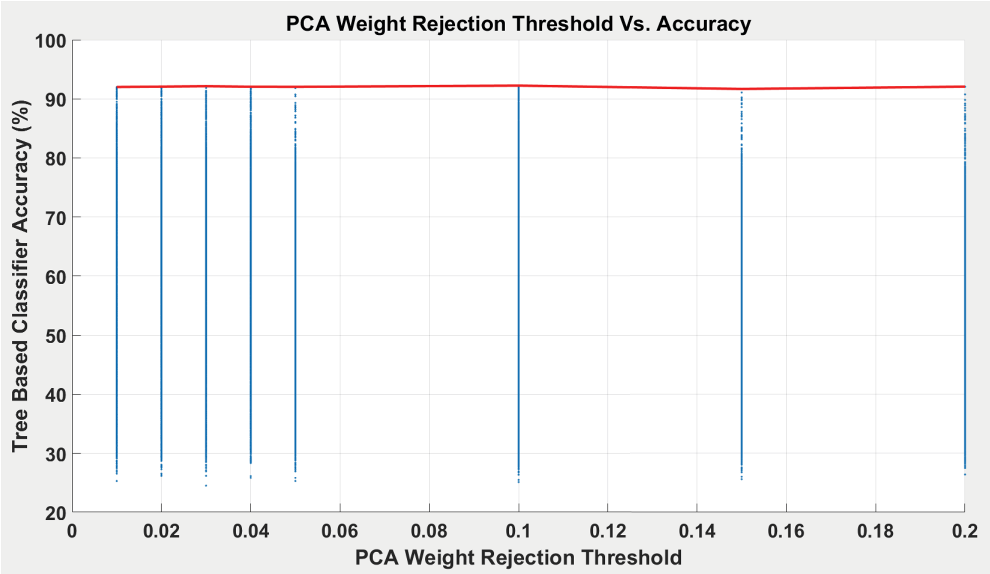


Figure 1 – PCA Weight Rejection Threshold (β) Vs. All Classifiers' Accuracies

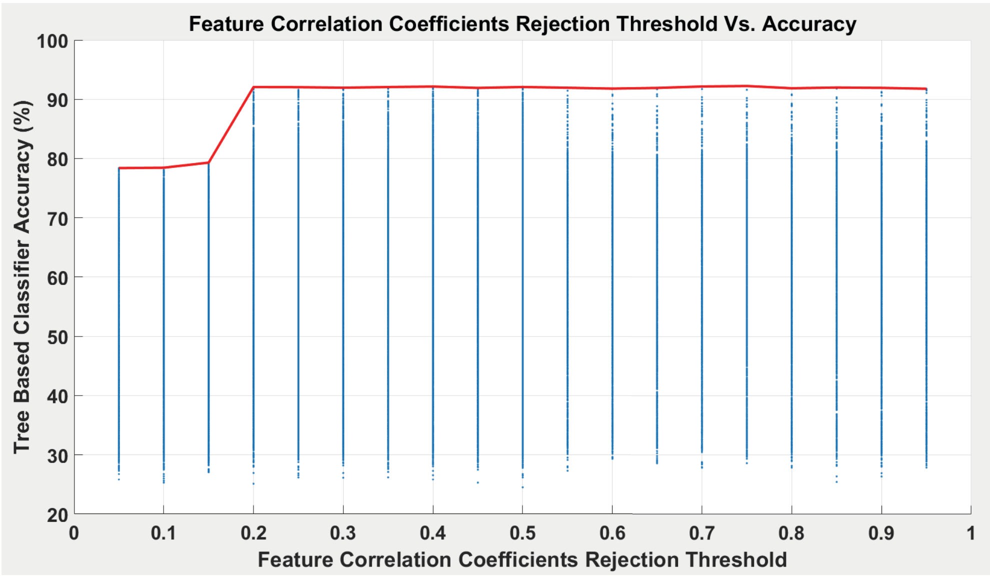


Figure 2 – Correlation Coefficient Rejection Threshold (α) Vs. All Classifiers' Accuracies – For 1 to 30 Features

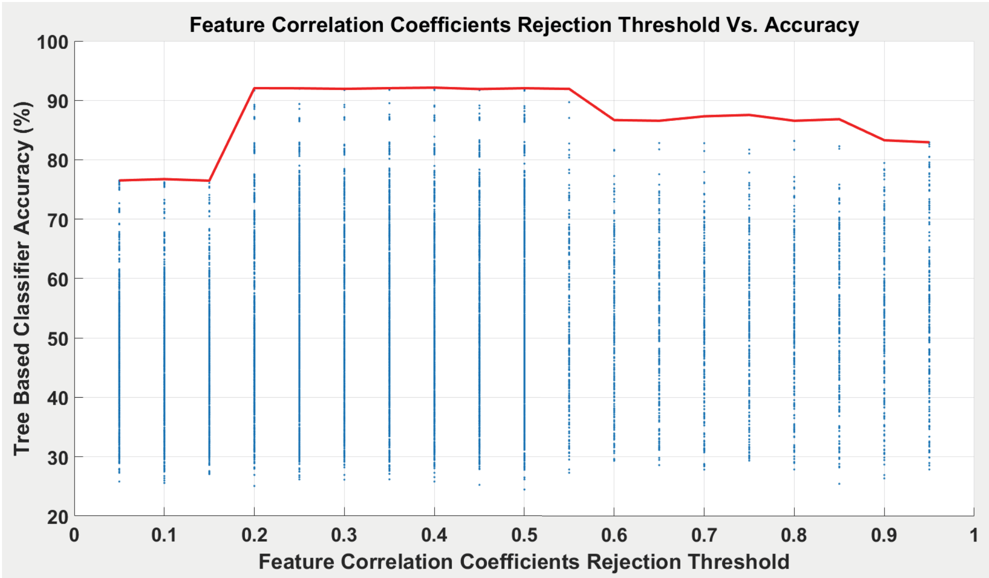


Figure 3 – Correlation Coefficient Rejection Threshold (α) Vs. All Classifiers' Accuracies – For 1 to 4 Features

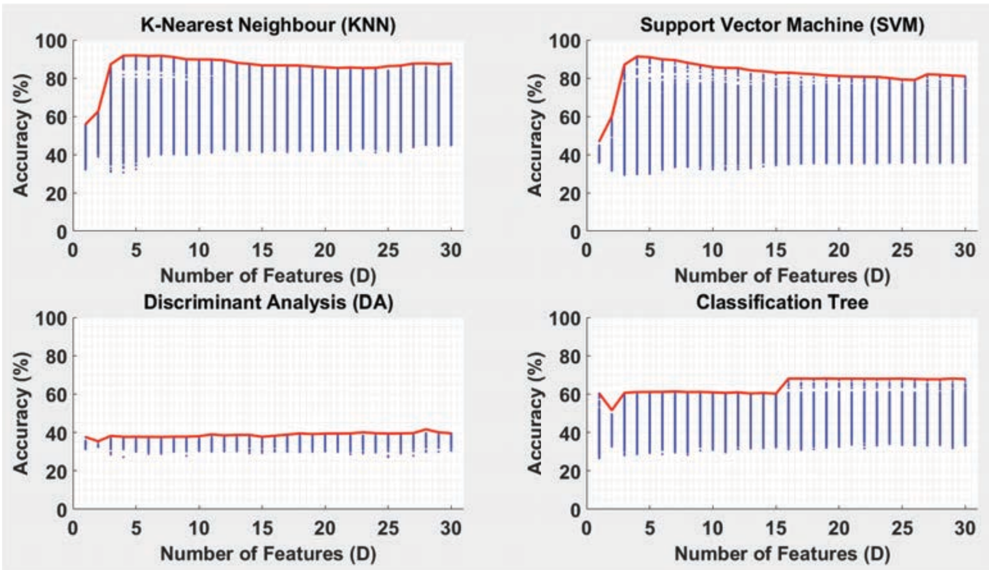


Figure 4 – Classifiers Performance Respect to Different Number of Features, Employed to Perform The Classification Process

Figure 4 presents the performance of classifiers according to different number of features. As it can be obtained by the graphs, the DA performance has not been changed considerably, by employing more or less features. The Classification Tree accuracy, on the other hand, has been increased by 10% after employing more than 15 features. The classification accuracy of both KNN and SVM classifiers, respect to the number of employed features, have almost the same behaviour. The optimal number of features for both KNN and SVM classifiers, to perform with highest accuracy, is four features.

2.2. Windowing Settings Evaluation

As it was discussed in Chapter 5, there are two tuning parameters for the windowing process; window type (Hamming vs. Tukey) and length (fixed vs. relative). As shown in Figure 5, the performance of KNN, SVM and Classification Tree is slightly better, while using Hamming window, compared to the Tukey window. Only the DA classifier performed slightly better, by employing the Tukey window.

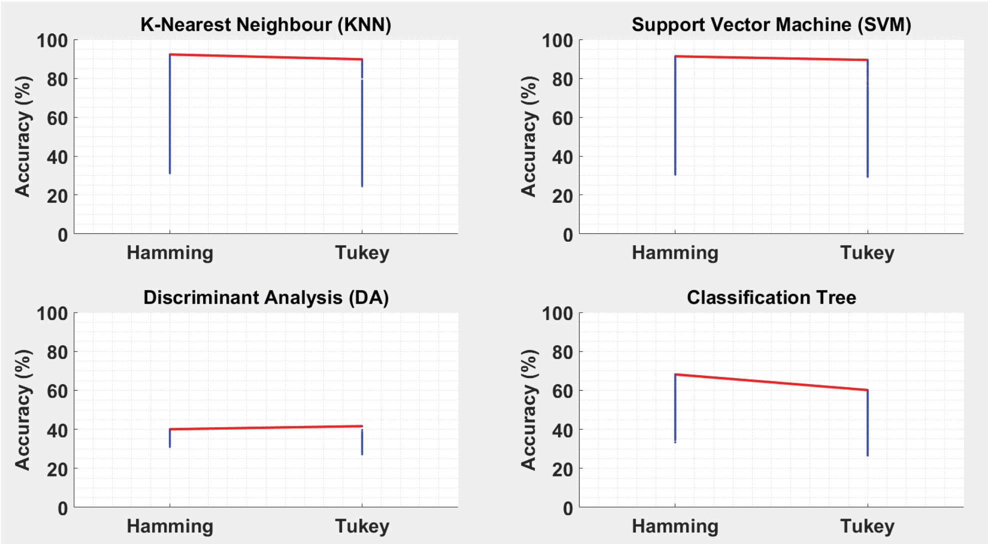


Figure 5 – Window Type Vs. Classifiers Accuracy

Figure 6 presents the classification accuracies of the classifiers, respect to different **fixed** window lengths. The maximum accuracy of the KNN is by employing the 10 seconds windowing process; although, the accuracy of the 5, 3 and 4 seconds windows are very similar to the 10 seconds window (less than or equal 1% difference). The Classification Tree performs best, by employing the 2 seconds windowing technique. The maximum accuracy of the SVM is with the 5 seconds windows, while the accuracy of the 3 and 4 seconds windows are very similar to the 5 seconds window (less than or equal 1.5% difference). The DA accuracy is less dependant to the window length, and its best performance is by employing the 55 seconds windowing algorithm. Figure 7 presents the classifiers performance, respect to different **relative** window lengths. Except DA classifiers, all classification algorithms perform better with **relatively** smaller windows, and achieve their maximum accuracy with window length equal to 5% of the stimuli duration. On the contrary, the DA classifier performs slightly better with relatively longer windows, and achieves its maximum classification accuracy with window length equal to the stimuli length.

Overall the performances of the KNN and SVM classifiers are better in **relative** windowing technique (compared to fixed). The performance of the DA classifier, using the fixed windowing technique, is not significantly different from the relative window lengths. Whereas the performance of the Classification Tree does not vary significantly, while using fixed (except 2 seconds window) or relative windowing techniques.

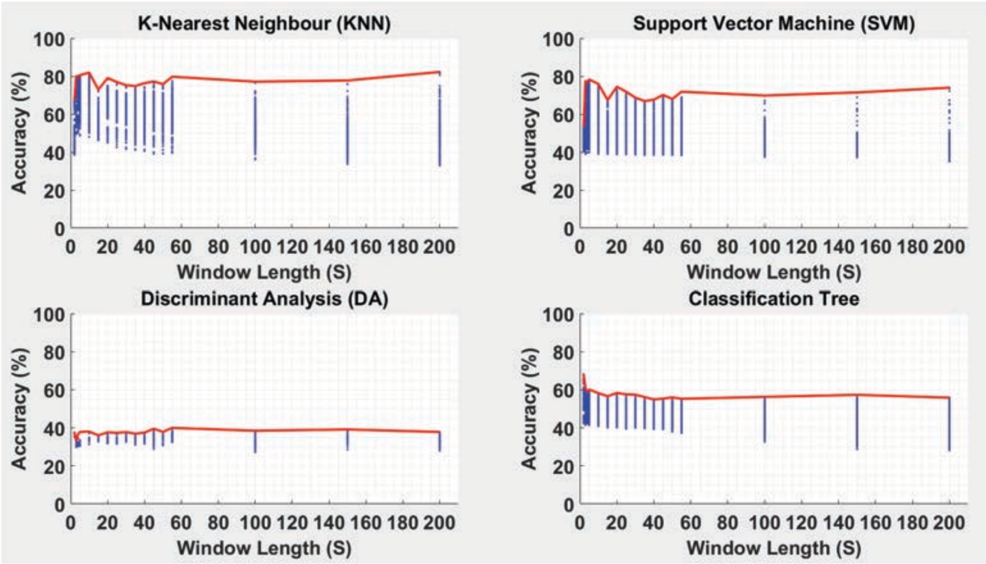


Figure 6 – Window Fixed Length Vs. Classifiers Accuracy

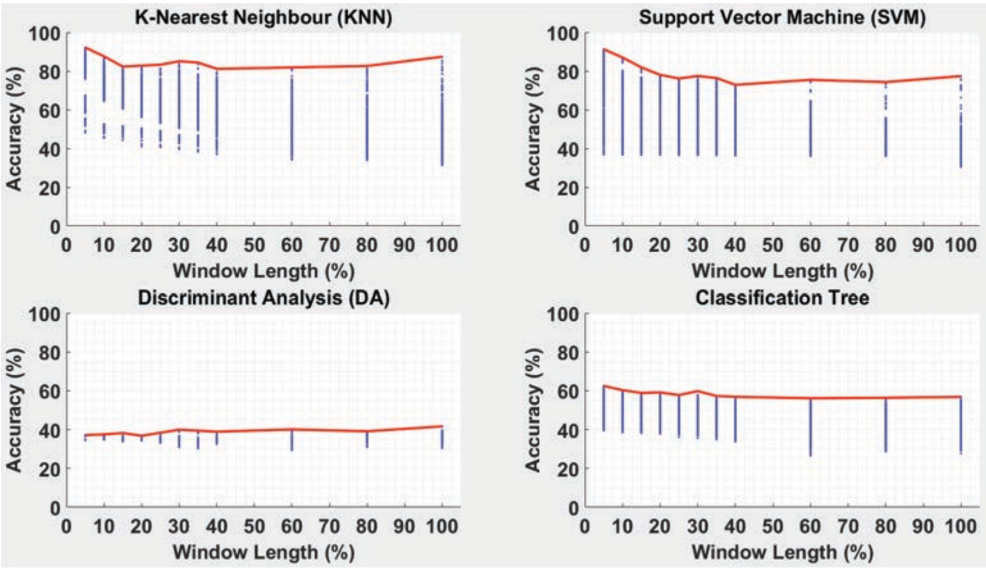


Figure 7 – Window Relative Length Vs. Classifiers Accuracy

3. Discussion

3.1. Most Optimal Features

As discussed is Section 2.2, the Hamming window outperforms the Tukey technique in majority of the classification techniques. Therefore, we decided to consider the Hamming technique, to identify the most optimal features. As a result, 28 different sets of the most optimal features[4] were extracted and the *unique features* (66 unique features), which are present in at least one of the 28 sets, were identified. To be able to sort unique features, according to their importance level, a ranking algorithm has been implemented. In this ranking algorithm, each of these 66 features is ranked with a value between 0 and 10. Equation 5 presents the feature ranking formula. Those, which ranked closer to 10, have appeared more frequently as the most important features (1st or 2nd) in different window lengths. Whereas, those, which ranked closer to zero, have appeared rarely as the less important features (30th, 29th, etc.) within a few window lengths. Table 1 presents these features, sorted according to their ranks. The rankings have not been used in the classification process, and have only been developed for appropriate presentation purposes.

$$Raw\_Rank(f) = \sum_{i\ for\ all\ 28\ windows} (31 - Win_i\_Ind(f))$$

$$Normalised\_Rank(f) = \frac{Raw\_Rank(f) - 1}{(28 \times 30) - 1} \times 10$$

$$Final\_Rank(f) = Normalised\_Rank(f) \times \frac{Num\_Wins\_Presence(f)}{28}$$

Equation 5 – Feature Rank Formula – 28 presents the 28 different window length and 30 presents the 30 most important features, that are extracted from each window length – $Win_i\_Ind(f)$ presents the importance level of a particular feature. E.g. if f_1 is the most relevant affective feature in window 1 and the 13th most relevant affective feature in window 2, then $Win_1\_Ind(f_1) = 1$ and $Win_2\_Ind(f_1) = 13$ – $Num\_Wins\_Presence(f)$ counts the number of windows, which a particular feature is identified as one of the 30 most optimal features in.

Table 1 – 66 Most Relevant Features Identified From 28 Different Windowing Techniques – Each Row Presents One or a Set of Features, Sorted According to Their Relevance Rank – A-B Presents the Subtracted Pair of two Symmetric Electrodes A and B – A & B Presents the Non-Subtracted Pair of two Symmetric Electrodes A and B

Feature		Rank (0-10)
GSR Low Frequency Power Squared Summation		10
GSR Fluctuation Frequency		9.2
GSR Minimum		6.82
Heart Rate Spectral Power Ratio		5.46
Alpha-Beta RMS Ratio db	F7	4.63
GSR Mean of the First Derivative		4.11
Alpha-Beta RMS Ratio db	AF3-AF4	4.09
	AF3	3.87
	F7-F8	3.43
Heart Rate Maximum		3.21
Alpha-Beta RMS Ratio db	AF4	3.15
	T7	2.92
	F3-F4	2.82
	O1-O2	2.7

[4] The 30 most relative features in Hamming windows with for fixed (17) and relative (11) variations.

	F8	P7	2.48	2.33
	F3	FC5	2.31	2.19
	O1		1.99	
Asymmetric Power RMS Ratio db in Alpha Rhythms	T7 & T8		1.84	
FC5-FC6 Channels Pair EEG_W (Power Summation)			1.58	
Alpha-Beta RMS Ratio db	O2	FC6	1.52	1.49
Alpha-Beta RMS Ratio db	F4		1.42	
Asymmetric Power RMS Ratio db Alpha Rhythms	AF3 & AF4		0.88	
Heart Rate Minimum			0.64	
Asymmetric Power RMS Ratio db Slow-Alpha Rhythms	T7 & T8		0.56	
Alpha-Beta RMS Ratio db	T8	T7-T8	0.55	0.43
Asymmetric Power RMS Ratio db Slow-Alpha Rhythms	AF3 & AF4		0.42	
Asymmetric Power RMS Ratio db Alpha Rhythms	FC5 & FC6		0.42	
Asymmetric Power RMS Ratio db Slow-Alpha Rhythms	P7 & P8		0.27	
Alpha-Beta RMS Ratio db	P8		0.24	
Asymmetric Power RMS Ratio db Slow-Alpha Rhythms	FC5 & FC6		0.24	
Asymmetric Power RMS Ratio db Alpha Rhythms	F3 & F4	P7 & P8	0.24	0.21
Asymmetric Power RMS Ratio db Slow-Alpha Rhythms	F7 & F8		0.12	
Alpha-Beta RMS Ratio db	P7-P8		0.1	
Asymmetric Power RMS Ratio db Slow-Alpha Rhythms	O1 & O2		0.1	
Gamma Power Summation	AF3-AF4		0.07	
Beta Power Summation	F7-F8	AF3-AF4	0.06	0.05
Asymmetric Power RMS Ratio db Alpha Rhythms	O1 & O2		0.04	
Beta Power Summation	T7		0.04	
Gamma Power RMS Ratio db	AF3		0.04	
Asymmetric Power RMS Ratio db Alpha Rhythms	F7 & F8		0.03	
Gamma Power Summation	FC5		0.03	
Beta Power Summation	FC5		0.02	
Gamma Power Summation	T7	FC6	0.01	0.01
Gamma Power RMS Ratio db	O1-O2		0.01	
Gamma Power RMS Ratio db	O2	F3-F4	<0.01	<0.01
Gamma Power RMS Ratio db	P7	T7	<0.01	<0.01
Gamma Power RMS Ratio db	T7-T8		<0.01	
Gamma Power RMS Ratio db	AF3-AF4	F7	<0.01	<0.01
Asymmetric Power RMS Ratio db Slow-Alpha Rhythms	F3 & F4		<0.01	
Beta Power Summation	AF4	F8	<0.01	<0.01
Beta Power Summation	FC5-FC6		<0.01	
Gamma Power Summation	FC5-FC6	F7-F8	<0.01	<0.01
Beta Power RMS Ratio db	P7-P8		<0.01	

As it can be seen in Table 1, the five most relative features are heart rate and GSR related measurements; whereas, the EEG related features are identified as less relevant features for affective states classification. The most important EEG-based features are the *Alpha-Beta* ratios, measured from all single and paired channels. Moreover, the EEG_W feature measured from the FC5-FC6 channels pair is identified as another considerably important EEG feature, for affective classification. Furthermore, the asymmetric power ratio measured in both *Alpha* and *Slow-Alpha* powers, from all single and paired channels, are another important EEG-based features. The Beta and Gamma brain rhythms have also been identified, with significantly low rankings, as less important features, related to affective states. On the other hand, the table suggests that, the Summation and RMS Ratio db algorithms (Chapter 5) have proved to be the most reliable methods to extract spectral powers.

3.2. Best Performing Classifiers

To be bale to compare the performance of all classification techniques, the best performing classifier (with the highest accuracy), for each window length, has been identified (Section 2.2). For all classifiers, α and β (for the feature selection algorithm) have been fixed to 0.2 and 0.1, respectively (Section 2.1). Also only the Hamming window technique has been investigated, due to its higher accuracy (Section 2.2). Therefore, the most optimal classifier setting (e.g. K value in KNN, etc.) and the most optimal number of features, within any window length, which results in highest classification accuracy, have been identified. As a result, 28 settings for each classification technique (KNN, SVM, DA and classification tree) have been identified. Figure 8 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)[5] showed that relative windowing technique is significantly different from fixed windowing algorithm ($P<0.001$). Moreover, the performances of different classifiers are significantly different, in terms of their classification accuracy ($P<0.001$). In average relative windowing technique, with a mean accuracy (across all classifiers) of 65.36%, slightly outperformed the fixed algorithm, with a mean accuracy (across all classifiers) of 61.13%. On the other hand, KNN (81.54% mean accuracy, across all window lengths) and SVM (75.13% mean accuracy, across all window lengths) perform better than classification tree (58.09% mean accuracy, across all window lengths). The DA classifier performs worst than the other three, with 38.22% mean accuracy, across all window lengths.

[5] Classifiers accuracy is considered as the dependent variables, while relative vs. fixed windowing technique and different classifiers as the independent parameters.

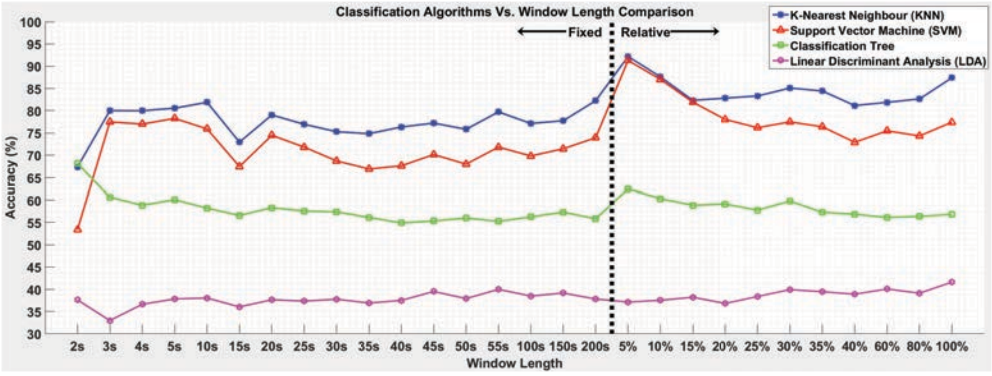


Figure 8 – Classification Methods Comparison – The Horizontal Axes Presents 28 Different Window Lengths; 17 Fixed (Left Side of the Vertical Dashed Line) and 11 Relative (Right Side of the Vertical Dashed Line)

3.3. Four Affective Clusters vs. Eight Emotion Labels Classification

Figure 9 presents the performance of KNN and SVM classifiers, while classifying the features space, respect to both four Affective Clusters and eight Emotion Labels. An Analysis of Variance (ANOVA)[6] showed that classification according to four Affective Clusters is significantly different from eight Emotion Labels (P=0.001). Moreover, the performances of KNN and SVM classifiers are significantly different, in terms of their classification accuracy (P<0.001). In average KNN (79.28% mean accuracy across different windowing techniques respect to both four and eight clusters) outperformed the SVM (71.96% mean accuracy across different windowing techniques respect to both four and eight clusters) algorithm. On the other hand, classification respect to four Affective Clusters (77.53% mean accuracy across all windowing and both classification techniques) performed slightly better than eight Emotion Labels (73.71% mean accuracy across all windowing and both classification techniques). Figure 10 presents the KNN and SVM classification F1-Score, averaged across classes[7]. No game in the experiment has been able to evoke sadness on the part of the participants; therefore, the classifiers, trained according to Emotion Labels, have not been able to classify any part of the features space, into "Sad" cluster.

To be able to compare the performance of the classifiers, according to their F1-Score, in each Affective Cluster or Emotion Label, an Analysis of Variance (ANOVA)[8] has been conducted. The analysis highlighted a significant difference in F1-Score generated by different classifiers and classification according to either Affective Clusters or Emotion Labels. Table 2 presents the mean F1-Scores for Affective Clusters, Emotion Labels and classifiers. In average the classification according to four Affective Clusters performed better, compared to eight Emotion Labels.

[6] Classifiers accuracy is considered as the dependent variables, while different classifiers and 4 affective clusters vs. 8-emotion labels classification technique as the independent parameters.
[7] F1-Score can be calculated within each class. Therefore in each windowing technique, 4 and 8 F1-Score for each (respectively) Affective Cluster, and Emotion Label, are calculated.
[8] Classifiers F1-Score is considered as the dependent variables, while four Affective Clusters vs. eight Emotion Labels classification and different classifiers as the independent parameters.

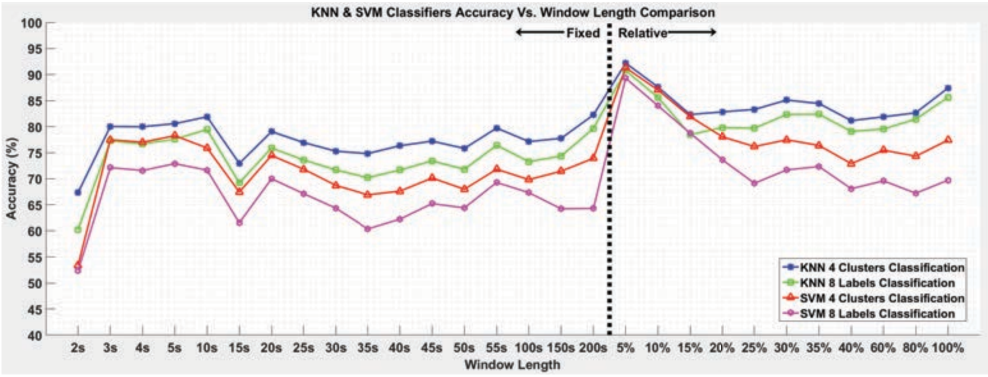


Figure 9 – Accuracy Comparison of KNN and SVM Classifiers, Considering Both 4 Affective Clusters and 8 Emotion Labels Classifications

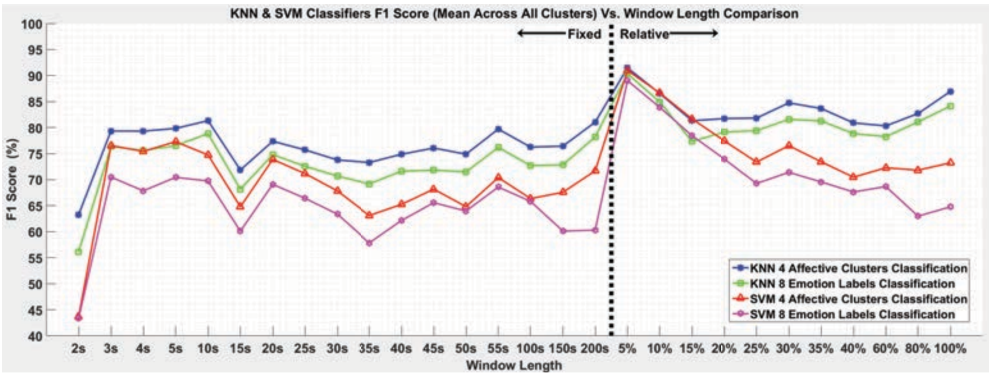


Figure 10 – KNN and SVM Mean F1-Score Across All Clusters

Table 2 – Mean F1-Scores Across Classifiers and All Windowing Techniques – (A - B) Presents the (A) 25^{th} and (B) 75^{th} Percentile

Emotion Label	**Mean F1-Score[9] (25^{th} – 75^{th} Percentiles)**	**Affective Cluster**	**Mean F1-Score[9] (25^{th} – 75^{th} Percentiles)**
Relaxed	73.55% (69.13% - 78.24%)	PVLA	79.77% (75.64% - 83.42%)
Content	74.05% (68.93% - 78.01%)		
Happy	68.95% (62.75% - 76.58%)	PVHPA	76.43% (72.73% - 80.74%)
Excited	75.50% (71.67% - 79.54%)		
Angry	73.15% (66.77% - 79.22%)	NVPA	76.10% (74.00% - 80.97%)
Afraid	66.04% (59.34% - 74.58%)		
Sad	Not Available	NVNA	71.38% (63.62% - 78.54%)
Bored	75.51% (72.28% - 79.85%)		
Emotion Labels Classifier	**Mean F1-Score[10] (25^{th} – 75^{th} Percentiles)**	**Affective Clusters Classifier**	**Mean F1-Score[10] (25^{th} – 75^{th} Percentiles)**
KNN	76.99% (73.63% - 80.49%)	KNN	79.85% (76.78% - 83.57%)
SVM	67.79% (63.06% -73.66%)	SVM	71.99% (68.13% - 77.29%)

9 Across different window lengths and classifiers.
10 Across different window lengths and classes.

References

Bergstra, J. & Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13(1), pp.281-305.

Cadima, J. & Jolliffe, I.T., 1995. Loadings and correlations in the interpretation of principal components Downloaded. *Journal of Applied Statistics*, 22(2), pp.203-14.

Hommel, G., 1983. Tests of the overall hypothesis for arbitrary dependence structures. *Biometrical Journal*, 25(5), pp.423–30.

Murphy, K.P., 2012. Estimation the Risk Using Cross Validation. In *Machine Learning: A Probabilistic Perspective*. MIT Press. p.209.

Murphy, K.P., 2012. Introduction. In *Machine Learning: A Probabilistic Perspective*. MIT Press. pp.2-12.

Murphy, K.P., 2012. Principle Component Analysis (PCA). In *Machine Learning A Probabilistic Perspective*. USA: MIT Press. pp.389-400.

Appendix L

Table 1 to 4 present the 112 best performing KNN and SVM classifiers' settings, according to the Affective Clusters and Emotion Labels. The "Number of Required Features" column presents the number of affective features, which are employed by the classifiers to conduct the classification process. Moreover, the "Features Index Number" column represent the features indices, within the feature matrix (refer to Appendix J for the features indices), which are employed by the classifiers to conduct the classification process.

Table 1 – The KNN Classifiers' Settings According to the Affective Clusters

Kth Nearest Neighbour (KNN)						
Classification According to Affective Clusters (PVLA, PVHPA, NVPA and NVNA)						
Window Type	Window Length	Number of Neighbours	Classification Accuracy	Classification F1-Score	Number of Required Features	Features Index Number
Hamming	2 Seconds	1	98.61%	98.53%	29	2, 3, 226, 237, 314, 323, 325, 377, 430, 440, 451, 532, 546, 556, 584, 585, 589, 590, 609, 630, 643, 648, 661, 670, 684, 729, 750, 751, 762
Hamming	3 Seconds	1	97.52%	97.42%	30	3, 4, 70, 203, 208, 216, 261, 300, 323, 336, 401, 406, 409, 410, 445, 447, 532, 559, 561, 585, 595, 630, 634, 648, 657, 679, 729, 750, 751, 763
Hamming	4 Seconds	1	98.28%	98.38%	30	3, 4, 57, 201, 203, 216, 237, 319, 324, 372, 401, 409, 429, 448, 474, 532, 561, 579, 613, 619, 630, 648, 657, 669, 687, 728, 729, 751, 762, 763
Hamming	5 Seconds	1	98.17%	98.07%	30	3, 4, 41, 57, 203, 211, 256, 326, 328, 400, 401, 441, 447, 448, 532, 561, 586, 591, 604, 609, 613, 630, 634, 652, 661, 670, 729, 750, 751, 763
Hamming	10 Seconds	1	97.39%	97.32%	30	3, 4, 41, 47, 57, 143, 201, 203, 316, 319, 323, 372, 406, 411, 441, 453, 532, 549, 561, 576, 589, 613, 634, 652, 657, 687, 729, 750, 751, 763
Hamming	15 Seconds	1	97.59%	97.53%	27	3, 4, 41, 47, 57, 211, 213, 298, 313, 319, 406, 409, 453, 522, 561, 576, 589, 604, 630, 652, 657, 670, 676, 687, 729, 751, 763
Hamming	20 Seconds	1	97.55%	97.28%	30	3, 4, 41, 47, 57, 211, 215, 219, 298, 319, 323, 406, 409, 429, 453, 532, 561, 576, 589, 604, 634, 652, 660, 666, 676, 687, 729, 750, 751, 763
Hamming	25 Seconds	1	97.11%	96.86%	30	3, 4, 41, 47, 57, 153, 215, 219, 298, 313, 319, 406, 409, 429, 453, 522, 561, 586, 589, 593, 613, 634, 639, 652, 656, 670, 676, 729, 750, 751
Hamming	30 Seconds	1	96.93%	96.06%	30	3, 41, 47, 57, 153, 215, 219, 298, 319, 395, 401, 406, 409, 429, 453, 522, 561, 586, 589, 591, 613, 630, 643, 651, 656, 670, 676, 729, 750, 751
Hamming	35 Seconds	1	96.70%	96.53%	30	3, 41, 47, 74, 153, 211, 215, 288, 319, 337, 395, 406, 409, 430, 453, 522, 561, 586, 589, 593, 613, 634, 643, 651, 660, 669, 676, 729, 750, 751
Hamming	40 Seconds	1	97.02%	96.55%	30	3, 47, 74, 153, 211, 215, 288, 319, 337, 406, 408, 433, 439, 453, 532, 561, 576, 579, 591, 613, 634, 643, 651, 656, 670, 676, 729, 742, 750, 751
Hamming	45 Seconds	1	97.08%	96.78%	29	3, 47, 74, 143, 201, 215, 288, 319, 337, 406, 409, 429, 453, 522, 561, 586, 591, 598, 604, 634, 638, 648, 660, 665, 676, 729, 742, 750, 751
Hamming	50 Seconds	1	96.60%	95.99%	28	3, 47, 74, 153, 201, 215, 288, 319, 326, 406, 409, 429, 453, 522, 561, 586, 591, 598, 613, 630, 643, 652, 660, 670, 729, 742, 750, 751
Hamming	55 Seconds	1	96.31%	95.60%	28	3, 47, 54, 71, 143, 216, 304, 319, 337, 406, 409, 429, 453, 522, 530, 561, 576, 592, 598, 614, 629, 643, 652, 656, 670, 729, 750, 751
Hamming	100 Seconds	1	94.55%	94.47%	28	3, 54, 211, 216, 319, 337, 407, 416, 429, 453, 482, 518, 530, 565, 566, 586, 591, 598, 604, 629, 639, 647, 670, 676, 729, 750, 751, 763
Hamming	150 Seconds	1	93.76%	92.36%	24	3, 64, 159, 216, 304, 329, 411, 430, 437, 453, 482, 526, 530, 566, 576, 598, 610, 643, 647, 650, 670, 742, 751, 763
Hamming	200 Seconds	1	94.20%	93.65%	30	3, 52, 64, 143, 201, 203, 325, 329, 397, 411, 418, 427, 430, 432, 453, 478, 507, 526, 561, 565, 576, 592, 598, 603, 643, 651, 669, 750, 751, 763
Hamming	5%	1	98.77%	96.74%	30	3, 4, 54, 203, 215, 226, 227, 287, 319, 323, 406, 429, 522, 561, 575, 576, 589, 603, 634, 638, 652, 660, 666, 683, 729, 748, 750, 751, 762, 763
Hamming	10%	1	97.63%	97.48%	27	3, 4, 64, 213, 215, 226, 298, 319, 323, 429, 453, 532, 561, 576, 589, 598, 634, 643, 651, 660, 666, 687, 729, 748, 750, 751, 763
Hamming	15%	1	97.87%	97.69%	27	3, 4, 64, 203, 211, 298, 325, 411, 429, 453, 532, 561, 597, 598, 603, 633, 642, 651, 660, 670, 676, 687, 729, 748, 750, 751, 763
Hamming	20%	1	97.64%	96.88%	25	3, 64, 153, 329, 406, 411, 429, 453, 530, 532, 561, 589, 598, 630, 643, 651, 660, 670, 676, 683, 729, 748, 750, 751, 763
Hamming	25%	1	97.83%	97.51%	22	3, 54, 319, 406, 429, 453, 522, 530, 561, 589, 592, 603, 634, 639, 651, 656, 670, 683, 748, 750, 751, 763
Hamming	30%	1	96.80%	96.46%	21	3, 216, 319, 411, 429, 453, 530, 561, 565, 589, 592, 634, 638, 651, 660, 666, 684, 748, 750, 751, 763
Hamming	35%	1	97.61%	97.13%	30	3, 64, 153, 211, 304, 315, 411, 426, 429, 453, 522, 549, 561, 579, 592, 598, 600, 634, 643, 651, 656, 658, 670, 676, 684, 742, 748, 750, 751, 763
Hamming	40%	1	96.31%	95.96%	30	3, 64, 211, 230, 319, 337, 381, 411, 424, 428, 429, 453, 522, 565, 566, 589, 592, 598, 603, 633, 643, 650, 651, 660, 670, 742, 748, 750, 751, 763
Hamming	60%	1	95.34%	93.67%	25	3, 52, 71, 159, 336, 379, 411, 417, 428, 453, 530, 561, 586, 592, 598, 638, 651, 658, 660, 669, 672, 745, 750, 751, 763
Hamming	80%	1	93.89%	92.70%	27	3, 53, 69, 201, 256, 319, 380, 407, 418, 424, 427, 453, 482, 523, 565, 576, 592, 598, 612, 639, 650, 651, 658, 745, 748, 751, 763
Hamming	100%	1	94.67%	92.66%	29	3, 26, 52, 54, 201, 325, 380, 397, 418, 424, 428, 439, 453, 523, 546, 561, 565, 586, 592, 598, 611, 642, 651, 654, 670, 748, 750, 751, 763

Table 2 – The KNN Classifiers' Settings According to the Emotion Labels

Kth Nearest Neighbour (KNN)						
Classification According to Emotion Labels (Relaxed, Content, Happy, Excited, Angry, Afraid and Bored - Sadness was not evoked)						
Window Type	**Window Length**	**Number of Neighbours**	**Classification Accuracy**	**Classification F1-Score**	**Number of Required Features**	**Features Index Number**
Hamming	2 Seconds	1	98.61%	98.51%	28	3, 4, 227, 260, 323, 399, 440, 447, 451, 527, 549, 556, 604, 609, 613, 630, 642, 643, 648, 661, 670, 688, 729, 750, 751, 759, 762, 763
Hamming	3 Seconds	1	98.35%	98.27%	30	3, 4, 213, 216, 315, 324, 331, 372, 406, 409, 410, 443, 445, 546, 559, 561, 579, 604, 613, 634, 643, 648, 670, 679, 688, 729, 750, 751, 759, 763
Hamming	4 Seconds	1	98.60%	98.54%	29	2, 3, 4, 57, 201, 203, 324, 325, 331, 406, 409, 410, 443, 445, 561, 586, 613, 634, 643, 648, 660, 669, 687, 728, 729, 750, 751, 762, 763
Hamming	5 Seconds	1	98.30%	98.21%	29	2, 3, 203, 324, 331, 336, 398, 400, 416, 433, 440, 445, 453, 561, 586, 604, 619, 634, 643, 648, 660, 670, 688, 729, 731, 750, 751, 762, 763
Hamming	10 Seconds	1	98.63%	98.67%	26	3, 4, 201, 325, 331, 406, 410, 433, 445, 532, 558, 561, 576, 592, 604, 634, 639, 652, 656, 670, 688, 727, 729, 750, 751, 762
Hamming	15 Seconds	1	98.34%	98.17%	29	2, 3, 47, 143, 325, 331, 337, 406, 410, 433, 445, 502, 522, 530, 548, 561, 589, 592, 604, 634, 643, 647, 660, 670, 687, 729, 731, 750, 751
Hamming	20 Seconds	1	97.83%	97.81%	27	3, 75, 215, 319, 327, 337, 410, 416, 443, 445, 522, 530, 558, 561, 586, 598, 604, 634, 643, 651, 660, 666, 688, 727, 729, 750, 751
Hamming	25 Seconds	1	97.25%	97.12%	27	3, 75, 215, 319, 327, 337, 409, 416, 433, 435, 522, 530, 548, 561, 586, 598, 613, 634, 639, 651, 656, 666, 688, 729, 731, 750, 751
Hamming	30 Seconds	1	97.08%	96.81%	30	3, 75, 215, 279, 319, 331, 337, 408, 416, 433, 445, 522, 530, 561, 564, 577, 586, 598, 613, 620, 633, 639, 651, 660, 670, 675, 684, 731, 750, 751
Hamming	35 Seconds	1	97.43%	97.35%	30	3, 74, 215, 317, 319, 337, 406, 408, 433, 435, 522, 530, 561, 562, 576, 577, 598, 614, 622, 633, 643, 651, 660, 675, 684, 729, 731, 750, 751, 755
Hamming	40 Seconds	1	96.91%	96.61%	24	3, 75, 215, 319, 337, 406, 408, 432, 522, 530, 566, 577, 586, 598, 622, 634, 643, 651, 660, 688, 731, 747, 750, 751
Hamming	45 Seconds	1	97.22%	96.86%	26	3, 57, 215, 317, 319, 337, 408, 433, 482, 518, 530, 561, 577, 586, 598, 604, 622, 633, 642, 651, 660, 686, 731, 742, 750, 751
Hamming	50 Seconds	1	96.93%	96.52%	27	3, 47, 74, 215, 288, 317, 319, 408, 432, 482, 518, 530, 566, 576, 587, 598, 614, 621, 634, 639, 651, 660, 686, 731, 742, 750, 751
Hamming	55 Seconds	1	97.73%	97.56%	30	3, 57, 65, 215, 317, 319, 408, 442, 453, 482, 518, 530, 566, 577, 586, 598, 604, 622, 633, 643, 651, 660, 668, 675, 686, 731, 742, 750, 751, 763
Hamming	100 Seconds	1	94.55%	94.29%	25	3, 40, 67, 211, 325, 326, 407, 418, 453, 478, 482, 491, 518, 525, 561, 565, 598, 614, 633, 639, 675, 686, 731, 746, 751
Hamming	150 Seconds	1	95.58%	95.28%	24	3, 47, 67, 120, 159, 201, 422, 432, 482, 491, 518, 525, 561, 576, 604, 630, 639, 674, 676, 685, 731, 742, 751, 763
Hamming	200 Seconds	1	95.05%	94.95%	28	3, 67, 167, 201, 295, 327, 422, 432, 453, 468, 478, 482, 501, 515, 565, 576, 604, 611, 634, 639, 640, 676, 679, 685, 727, 742, 748, 751
Hamming	5%	1	98.89%	98.72%	28	3, 287, 314, 325, 331, 398, 416, 430, 433, 445, 459, 561, 576, 602, 614, 634, 643, 648, 660, 666, 687, 729, 731, 748, 750, 751, 762, 763
Hamming	10%	1	97.79%	97.70%	30	3, 297, 313, 329, 331, 337, 416, 430, 432, 445, 459, 522, 561, 564, 586, 598, 602, 633, 639, 651, 660, 666, 683, 731, 742, 748, 750, 751, 762, 763
Hamming	15%	1	97.49%	97.47%	23	3, 215, 331, 337, 406, 430, 432, 459, 530, 561, 597, 602, 633, 639, 651, 660, 683, 727, 748, 750, 751, 762, 763
Hamming	20%	1	97.46%	97.27%	29	3, 256, 302, 314, 319, 327, 337, 406, 430, 432, 459, 500, 520, 522, 561, 596, 598, 603, 633, 639, 651, 660, 684, 731, 748, 750, 751, 762, 763
Hamming	25%	1	97.23%	96.96%	18	3, 319, 430, 432, 530, 561, 589, 622, 643, 651, 660, 683, 727, 742, 748, 750, 751, 763
Hamming	30%	1	96.93%	96.56%	28	3, 201, 319, 331, 362, 398, 421, 430, 442, 453, 478, 482, 491, 518, 561, 565, 589, 614, 634, 638, 651, 660, 684, 731, 748, 750, 751, 763
Hamming	35%	1	97.02%	96.88%	23	3, 211, 329, 381, 398, 430, 433, 453, 482, 501, 518, 561, 589, 621, 639, 647, 660, 731, 743, 748, 750, 751, 763
Hamming	40%	1	97.48%	97.74%	20	3, 201, 331, 398, 430, 432, 482, 491, 566, 576, 614, 621, 639, 647, 660, 731, 747, 748, 751, 763
Hamming	60%	1	97.30%	96.71%	30	3, 167, 201, 325, 423, 430, 431, 432, 468, 482, 491, 567, 576, 592, 600, 614, 633, 639, 651, 653, 656, 676, 679, 685, 727, 746, 748, 750, 751, 763
Hamming	80%	1	95.65%	94.44%	29	3, 59, 110, 201, 325, 379, 397, 418, 424, 453, 466, 482, 484, 491, 523, 566, 576, 614, 632, 638, 640, 654, 656, 676, 685, 745, 748, 751, 763
Hamming	100%	1	95.11%	92.32%	30	3, 59, 167, 201, 325, 418, 424, 430, 448, 453, 466, 478, 482, 491, 515, 565, 569, 591, 604, 632, 640, 642, 654, 669, 676, 679, 685, 748, 751, 763

Table 3 – The SVM Classifiers' Settings According to the Affective Clusters

Supprt Vector Machine (SVM)							
Classification According to Affective Clusters (PVLA, PVHPA, NVPA and NVNA)							
Window Type	Window Length	Kernel Function	Kernel Scale	Classification Accuracy	Classification F1-Score	Number of Required Features	Features Index Number
Hamming	2 Seconds	Gaussian	2	98.07%	97.78%	26	2, 3, 226, 237, 314, 323, 325, 377, 430, 440, 451, 532, 546, 556, 584, 585, 589, 590, 609, 630, 643, 648, 661, 670, 684, 729, 750, 751, 762
Hamming	3 Seconds	Gaussian	2	96.76%	96.57%	27	3, 4, 70, 203, 208, 261, 300, 323, 336, 401, 406, 410, 445, 447, 532, 559, 585, 595, 630, 634, 648, 657, 679, 729, 750, 751, 763
Hamming	4 Seconds	Gaussian	2	97.39%	97.13%	30	3, 4, 57, 201, 203, 216, 237, 319, 324, 372, 401, 409, 429, 448, 474, 532, 561, 579, 613, 619, 630, 648, 657, 669, 687, 728, 729, 751, 762, 763
Hamming	5 Seconds	Gaussian	2	96.95%	96.77%	29	3, 4, 41, 57, 203, 211, 256, 326, 328, 400, 401, 441, 447, 532, 561, 586, 591, 604, 609, 613, 630, 634, 652, 661, 670, 729, 750, 751, 763
Hamming	10 Seconds	Gaussian	2	95.34%	95.30%	23	3, 4, 41, 57, 201, 203, 319, 323, 406, 411, 453, 532, 561, 576, 589, 613, 634, 652, 657, 729, 750, 751, 763
Hamming	15 Seconds	Gaussian	3	95.13%	94.61%	28	3, 4, 41, 47, 57, 211, 213, 298, 313, 319, 406, 409, 429, 453, 522, 561, 576, 589, 604, 630, 652, 657, 670, 676, 687, 729, 751, 763
Hamming	20 Seconds	Gaussian	3	94.55%	94.40%	30	3, 4, 41, 47, 57, 211, 215, 219, 298, 319, 323, 406, 409, 429, 453, 532, 561, 576, 589, 604, 634, 652, 660, 666, 676, 687, 729, 750, 751, 763
Hamming	25 Seconds	Gaussian	3	93.89%	93.39%	30	3, 4, 41, 47, 57, 153, 215, 219, 298, 313, 319, 406, 409, 429, 453, 522, 561, 586, 589, 593, 613, 634, 639, 652, 656, 670, 676, 729, 750, 751
Hamming	30 Seconds	Gaussian	2	92.24%	91.47%	19	3, 57, 215, 298, 319, 406, 409, 429, 453, 522, 561, 589, 630, 643, 651, 656, 676, 729, 751
Hamming	35 Seconds	Gaussian	2	93.32%	92.87%	22	3, 74, 153, 215, 288, 319, 406, 409, 430, 453, 522, 561, 586, 589, 634, 643, 651, 660, 676, 729, 750, 751
Hamming	40 Seconds	Gaussian	2	93.03%	92.33%	22	3, 74, 153, 215, 288, 319, 406, 408, 439, 453, 532, 561, 579, 613, 634, 643, 651, 656, 676, 729, 750, 751
Hamming	45 Seconds	Gaussian	2	93.08%	92.28%	24	3, 74, 143, 215, 288, 319, 406, 409, 429, 453, 522, 561, 586, 598, 604, 634, 638, 648, 660, 665, 729, 742, 750, 751
Hamming	50 Seconds	Gaussian	2	91.96%	91.14%	23	3, 74, 201, 288, 319, 406, 409, 429, 453, 522, 561, 586, 598, 613, 630, 643, 652, 660, 670, 729, 742, 750, 751
Hamming	55 Seconds	Gaussian	2	91.25%	89.64%	22	3, 71, 143, 216, 319, 337, 406, 429, 453, 522, 530, 561, 576, 598, 614, 629, 643, 652, 656, 729, 750, 751
Hamming	100 Seconds	Gaussian	3	87.81%	86.65%	30	3, 54, 71, 159, 211, 216, 319, 337, 407, 416, 429, 453, 482, 518, 530, 565, 566, 586, 591, 598, 604, 629, 639, 647, 670, 676, 729, 750, 751, 763
Hamming	150 Seconds	Gaussian	2	87.52%	84.13%	22	3, 64, 159, 216, 329, 411, 437, 453, 482, 526, 530, 566, 576, 598, 610, 643, 647, 650, 670, 742, 751, 763
Hamming	200 Seconds	Cubic	NA	86.48%	85.07%	28	3, 52, 64, 143, 201, 203, 325, 329, 397, 411, 418, 427, 432, 453, 478, 507, 561, 565, 576, 592, 598, 603, 643, 651, 669, 750, 751, 763
Hamming	5%	Gaussian	2	97.45%	97.06%	25	3, 4, 54, 203, 215, 226, 287, 319, 323, 429, 522, 561, 576, 589, 603, 634, 638, 652, 660, 666, 683, 729, 751, 762, 763
Hamming	10%	Gaussian	2	95.08%	94.29%	24	3, 4, 64, 213, 226, 298, 319, 323, 429, 453, 532, 561, 576, 598, 634, 643, 651, 660, 666, 687, 729, 748, 751, 763
Hamming	15%	Gaussian	3	94.65%	94.69%	30	3, 4, 64, 203, 211, 298, 325, 406, 411, 427, 429, 453, 532, 561, 586, 597, 598, 603, 633, 642, 651, 660, 670, 676, 687, 729, 748, 750, 751, 763
Hamming	20%	Gaussian	2	93.61%	92.56%	20	3, 64, 329, 411, 429, 453, 530, 532, 561, 589, 630, 643, 651, 660, 670, 683, 729, 750, 751, 763
Hamming	25%	Gaussian	2	93.85%	92.30%	22	3, 54, 319, 406, 429, 453, 522, 530, 561, 589, 592, 603, 634, 639, 651, 656, 670, 683, 748, 750, 751, 763
Hamming	30%	Gaussian	2	92.71%	91.61%	21	3, 216, 319, 411, 429, 453, 530, 561, 565, 589, 592, 634, 638, 651, 660, 666, 684, 748, 750, 751, 763
Hamming	35%	Gaussian	3	92.14%	90.57%	30	3, 64, 153, 211, 304, 315, 411, 426, 429, 453, 522, 549, 561, 579, 592, 598, 600, 634, 643, 651, 656, 658, 670, 676, 684, 742, 748, 750, 751, 763
Hamming	40%	Gaussian	3	90.95%	89.10%	29	3, 64, 211, 230, 319, 337, 381, 411, 424, 428, 429, 453, 522, 565, 566, 589, 592, 598, 633, 643, 650, 651, 660, 670, 742, 748, 750, 751, 763
Hamming	60%	Gaussian	2	88.56%	85.44%	25	3, 52, 71, 159, 336, 379, 411, 417, 428, 453, 530, 561, 586, 592, 598, 638, 651, 658, 660, 669, 672, 745, 750, 751, 763
Hamming	80%	Gaussian	3	86.02%	80.17%	30	3, 53, 69, 201, 256, 319, 325, 380, 407, 418, 424, 427, 453, 482, 523, 565, 576, 581, 592, 598, 612, 639, 650, 651, 658, 745, 748, 750, 751, 763
Hamming	100%	Gaussian	3	88.01%	83.96%	30	3, 26, 52, 54, 201, 325, 380, 397, 418, 424, 428, 439, 453, 466, 523, 546, 561, 565, 586, 592, 598, 611, 642, 651, 654, 670, 748, 750, 751, 763

Table 4 – The SVM Classifiers' Settings According to the Emotion Labels

Supprt Vector Machine (SVM)							
Classification According to Emotion Labels (Relaxed, Content, Happy, Excited, Angry, Afraid and Bored - Sadness was not evoked)							
Window Type	**Window Length**	**Kernel Function**	**Kernel Scale**	**Classification Accuracy**	**Classification F1-Score**	**Number of Required Features**	**Features Index Number**
Hamming	2 Seconds	Cubic	NA	98.30%	98.38%	28	3, 4, 227, 260, 323, 399, 440, 447, 451, 527, 549, 556, 604, 609, 613, 630, 642, 643, 648, 661, 670, 688, 729, 750, 751, 759, 762, 763
Hamming	3 Seconds	Gaussian	2	97.17%	97.18%	24	3, 4, 213, 315, 324, 406, 409, 410, 443, 445, 561, 579, 604, 613, 634, 643, 648, 670, 688, 729, 750, 751, 759, 763
Hamming	4 Seconds	Gaussian	3	97.78%	97.84%	30	2, 3, 4, 57, 201, 203, 324, 325, 331, 406, 409, 410, 440, 443, 445, 561, 586, 613, 634, 643, 648, 660, 669, 687, 728, 729, 750, 751, 762, 763
Hamming	5 Seconds	Gaussian	3	97.48%	97.52%	30	2, 3, 203, 324, 331, 336, 398, 400, 416, 433, 440, 445, 448, 453, 561, 586, 604, 619, 634, 643, 648, 660, 670, 688, 729, 731, 750, 751, 762, 763
Hamming	10 Seconds	Gaussian	2	97.33%	97.16%	26	3, 4, 201, 325, 331, 406, 410, 433, 445, 532, 558, 561, 576, 592, 604, 634, 639, 652, 656, 670, 688, 727, 729, 750, 751, 762
Hamming	15 Seconds	Gaussian	2	96.32%	96.14%	23	3, 47, 325, 331, 406, 410, 433, 445, 522, 530, 561, 589, 604, 634, 643, 647, 660, 670, 687, 729, 731, 750, 751
Hamming	20 Seconds	Gaussian	3	95.69%	95.50%	30	2, 3, 4, 47, 75, 215, 319, 327, 337, 410, 416, 443, 445, 522, 530, 558, 561, 586, 598, 604, 634, 643, 651, 660, 666, 688, 727, 729, 750, 751
Hamming	25 Seconds	Gaussian	3	94.44%	94.25%	30	3, 4, 75, 215, 319, 327, 337, 409, 416, 433, 435, 522, 530, 548, 561, 569, 586, 598, 613, 634, 639, 651, 656, 666, 675, 688, 729, 731, 750, 751
Hamming	30 Seconds	Gaussian	3	93.69%	93.24%	30	3, 75, 215, 279, 319, 331, 337, 408, 416, 433, 445, 522, 530, 561, 564, 577, 586, 598, 613, 620, 633, 639, 651, 660, 670, 675, 684, 731, 750, 751
Hamming	35 Seconds	Gaussian	3	93.76%	93.65%	30	3, 74, 215, 317, 319, 337, 406, 408, 433, 435, 522, 530, 561, 562, 576, 577, 598, 614, 622, 633, 643, 651, 660, 675, 684, 729, 731, 750, 751, 755
Hamming	40 Seconds	Gaussian	2	92.96%	92.45%	22	3, 75, 215, 319, 406, 408, 432, 522, 530, 566, 577, 586, 598, 634, 643, 651, 660, 688, 731, 747, 750, 751
Hamming	45 Seconds	Gaussian	2	92.23%	91.96%	18	3, 57, 215, 319, 408, 433, 530, 561, 577, 598, 622, 633, 642, 651, 731, 742, 750, 751
Hamming	50 Seconds	Gaussian	3	92.23%	91.10%	27	3, 47, 74, 215, 288, 317, 319, 408, 432, 482, 518, 530, 566, 576, 587, 598, 614, 621, 634, 639, 651, 660, 686, 731, 742, 750, 751
Hamming	55 Seconds	Gaussian	3	93.08%	93.23%	30	3, 57, 65, 215, 317, 319, 408, 442, 453, 482, 518, 530, 566, 577, 586, 598, 604, 622, 633, 643, 651, 660, 668, 675, 686, 731, 742, 750, 751, 763
Hamming	100 Seconds	Gaussian	3	87.94%	86.76%	30	3, 40, 67, 211, 325, 326, 362, 407, 418, 432, 453, 478, 482, 491, 518, 525, 530, 561, 565, 598, 614, 633, 639, 675, 686, 731, 746, 750, 751, 763
Hamming	150 Seconds	Gaussian	2	87.45%	85.73%	18	3, 67, 201, 422, 432, 482, 491, 525, 561, 576, 604, 630, 639, 676, 685, 731, 751, 763
Hamming	200 Seconds	Cubic	NA	85.05%	84.84%	28	3, 67, 167, 201, 295, 327, 422, 432, 453, 468, 478, 482, 501, 515, 565, 576, 604, 611, 634, 639, 640, 676, 679, 685, 727, 742, 748, 751
Hamming	5%	Gaussian	3	97.48%	97.40%	29	3, 236, 287, 314, 325, 331, 398, 416, 430, 433, 445, 459, 561, 576, 602, 614, 634, 643, 648, 660, 666, 687, 729, 731, 748, 750, 751, 762, 763
Hamming	10%	Gaussian	3	95.25%	95.06%	30	3, 297, 313, 329, 331, 337, 416, 430, 432, 445, 459, 522, 561, 564, 586, 598, 602, 633, 639, 651, 660, 666, 683, 731, 742, 748, 750, 751, 762, 763
Hamming	15%	Gaussian	2	94.23%	94.41%	17	3, 331, 406, 430, 432, 530, 561, 597, 633, 639, 651, 683, 727, 748, 750, 751, 763
Hamming	20%	Gaussian	2	93.50%	92.81%	20	3, 327, 337, 406, 430, 432, 459, 520, 561, 596, 633, 639, 651, 660, 684, 731, 748, 750, 751, 763
Hamming	25%	Gaussian	2	93.55%	93.32%	18	3, 319, 430, 432, 530, 561, 589, 622, 643, 651, 660, 683, 727, 742, 748, 750, 751, 763
Hamming	30%	Gaussian	2	91.59%	90.70%	20	3, 201, 319, 398, 421, 430, 442, 453, 491, 561, 589, 638, 651, 660, 684, 731, 748, 750, 751, 763
Hamming	35%	Cubic	NA	91.46%	91.02%	28	3, 211, 304, 329, 337, 381, 398, 422, 430, 433, 453, 482, 501, 518, 561, 589, 614, 621, 634, 639, 647, 660, 731, 743, 748, 750, 751, 763
Hamming	40%	Gaussian	3	90.95%	90.05%	30	3, 167, 201, 331, 398, 421, 430, 432, 453, 466, 478, 482, 491, 518, 566, 576, 598, 614, 621, 633, 639, 647, 660, 683, 731, 747, 748, 750, 751, 763
Hamming	60%	Gaussian	3	91.47%	89.80%	30	3, 167, 201, 325, 423, 430, 431, 432, 468, 482, 491, 567, 576, 592, 600, 614, 633, 639, 651, 653, 656, 676, 679, 685, 727, 746, 748, 750, 751, 763
Hamming	80%	Cubic	NA	85.56%	84.61%	30	3, 59, 110, 201, 325, 379, 397, 418, 424, 432, 453, 466, 482, 484, 491, 523, 566, 576, 614, 632, 638, 640, 654, 656, 676, 685, 745, 748, 751, 763
Hamming	100%	Cubic	NA	83.79%	81.15%	30	3, 59, 167, 201, 325, 418, 424, 430, 448, 453, 466, 478, 482, 491, 515, 565, 569, 591, 604, 632, 640, 642, 654, 669, 676, 679, 685, 748, 751, 763