Games 4 VRains: Affective Gaming for Working Memory Training in Virtual Reality

Daniel Gabana

PhD thesis

School of Electronic Engineering and Computer Science
Queen Mary University of London

2018

Abstract

The explosion of Virtual Reality (VR) in the last few years, thanks to the introduction of affordable Head-Mounted Displays (HMD), has increased the interest in this technology for research. One of the main research areas using VR is the field of cognitive and physical rehabilitation or training. Although it is in early stages, many researchers have shown the positive effects of the higher levels of immersion, often reported in VR, on cognitive skills. Video games have also been used for cognitive training due to their capacity to engage and motivate players. Recent findings have demonstrated that by adapting the game to the player's performance, real cognitive benefits can be achieved as the adaptation offers a personalised cognitive training program. However, this adaptation normally considers just performance metrics and ignores other crucial aspects like the player's affective states or experience. Arousal and valence have generally been shown to enhance the subjects' cognitive skills and thus should also be considered when adapting a game for cognitive training.

Following these findings, this thesis investigates the effects of affect and performance-based adaptation of a VR video game on player's working memory (WM) performance. An initial pilot study explores suitable ways of measuring player's arousal and valence levels through physiological and behavioural cues. In a second study, the effects of immersion, arousal and valence on player's WM performance in Desktop and VR gaming are examined. The results of this study show that players in an optimal affective state can significantly improve their WM performance, supporting the incorporation of affective metrics in the adaptation engine. Thus, an adaptation engine was developed, implemented and tested to automatically adjust the game's difficulty level depending on the player's performance and the detected affective state. Two machine learning algorithms in the adaptation engine recognise and classify player's arousal and valence levels using physiological and behavioural features for adaptive decision making. Across the three studies presented, this thesis makes the following novel contributions. It shows that, i) VR is a suitable medium for cognitive training since the elicited high levels of immersion have a positive effect on players' WM performance, ii) positive affective states help subjects to achieve a better WM performance, and ii) difficulty adaptation is more beneficial for subjects with low WM capacity. During this process, it also provides a new methodology for affect recognition in VR gaming and a novel adaptation engine compounded by affect and performance metrics. Therefore, this work proposes that game-based cognitive training would be improved by VR, especially by the use of

affective and performance metrics for dynamic adaption, resulting in a highly personalised and more effective training experience.

Acknowledgements

I would like to thank my primary supervisor, Dr. Laurissa Tokarchuk, for her guidance and support over the last four years. It has been a pleasure to work with you, and i wish you the best in your future. I would also like to thank to Dr. Hatice Gunes, my secondary supervisor during the my first year and external collaborator after she moved to the University of Cambridge. Thanks for all you valuable input in this research, i wish you the best with your son Mavi. I would like to extend my gratitude to Dr. Emily M Hannon for her wise advice on the cognitive research of this thesis. I hope you enjoy your well deserved break.

Special thanks to my mother Clara, my father Antonio and my brother Victor for all your patience and encouragement during these years. I am really looking forward to see Marcos growing up and enjoy those moment with all of you - Yasmina, this goes for you too. I am grateful for having a second Spanish family in London who also supported and encouraged me throughout these years. The list of names would be very long as there has been a lot of transit of people during these years, but i appreciate the friendship and love that every one of you have shared with me.

Thanks to the Media & Arts Technology program for giving me the opportunity to make this PhD in such a great and inspirational environment full of amazing creative people. I would like to thank in particular to Astrid Bin, Lida Theodorou, Sophie Skach, Lucia Marengo, Soomi Park-Wells, Delia Fano and Dave Moffat. It has been great meeting and working with you. You are great people and i wish you the best in your careers. Thanks also to the MAT and Queen Mary staff for their help and work that made my journey a bit easier: Dr. Nick Bryan-Kinns, Jonathan Winfield, Geetha Bommireddy, Edward Hoskins, Hayley Cork, Ho Kok Huen and Kelly Peake. Huge thanks as well to Tim Kay, Keith Bancroft and Tom King for their invaluable IT support and help in the most critical moments.

Finally, I also would like to thank to some good friends here in England (Ross, Jonny, Fede, Leah, Stephen, Nina, Susan, Mandy) and in Spain (Dani, Alex, Rafa, Lucas, Vija, Serch, Jose, Willy, Deviasa Crew...) for their kindness, friendship and support during my time in London.

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) as part of the Doctoral Training Centre in Media and Arts Technology at Queen Mary University of London (ref: EP/G03723X/1).

Statement of Originality

I, Daniel Gabana, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by oth- ers, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below. I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material. I accept that the College has the right to use plagia- rism detection software to check the electronic version of the thesis. I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date:

Contents

List of Figures

List of Tables

List of abbreviations

ADHD	Attention Deficit Hyperactivity Disorder
ANOVA	Analysis of Variance
AOST	Automated Operational Span Task
AVNN	Average of N-N Intervals
BPM	Beats Per Minute
CAVE	Cave Automatic Virtual Environment
CLVG	Closed-Loop Video Game
CPT	Continuous Performance Task
DDA	Dynamic Difficulty Adjustment
DF	Degrees of Freedom
ECG	Electrocardiogram
EDA	Electrodermal activity
EEG	Electroencephalography
EMG	Electromyograph
fMRI	Functional Magnetic Resonance Imaging
GEQ	Game Experience Questionnaire
GSR	Galvanic Skin Response
HMD	Head-Mounted Display
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Inter-Beat Interval
KNN	k-Nearest Neighbours
LM	Leap Motion
LME	Linear Mixed Effects
LF/HF	Low Frequency / High Frequency Ratio
OSpan	Operational Span Task
PC	Personal Computer
QoM	Quantity of Motion
RBF	Radial Basis Function
rMSSD	Root Mean Square of the Successive Differences
SD	Standard Deviation
SVM	Support Vector Machine
TBI	Traumatic Brain Injuries
VR	Virtual Reality
WM	Working Memory

Chapter 1

Introduction

Since the beginning of affective computing in 1995 by Rosalind Picard, who defined it as "*computing that relates to, arises from or deliberately influences emotions*" [121], this field of research has been applied to many areas such as robotics, social interactions and human-computer interaction (HCI). For example, it has been used to detect the emotions of computer users employing physiological or motion sensors in order to adapt and improve the interaction depending on how the user is feeling [156]. Its application to video games, called affective gaming, aims to recognise and influence the player's affective states, represented in the two dimensional space of valence and arousal [62]. This PhD thesis investigates the use of affective gaming on games for cognitive training, specifically for working memory (WM) training. Since arousal and valence can have important effects on a players' cognitive skills, affective gaming can bring interesting insights in the research of game-based cognitive training programs. Throughout the three studies, this research explores the recognition of player's affective states using physiological and behavioural cues. The detected affective states are then used for real-time adaptation of a custom-made Virtual Reality (VR) game called *Memory Break*. Furthermore, this research explores the suitability of VR for cognitive training.

This chapter is structured as follows: section 1.1 documents the motivations of this thesis in the context of previous research. Next, section 1.2 informs about the aims and approach of this research following the motivations described. The main contributions of this thesis are briefly outlined in section 1.3. Finally, sections 1.4 and 1.5 report the publications associated with this PhD research and the structure of this thesis.

1.1 Motivations

In the age of artificial intelligence and intelligent machines, computers are becoming more interactive. Interaction with technology has traditionally been unidirectional in that the user commands a passive computer that obeys orders [55]. A novel approach for making computers more interactive by including an adaptation engine has been proposed by many researchers [55, 99, 7, 24]. This adaptation engine consists of adapting the interaction depending on the user's actions to keep a bidirectional communication, making the computer aware of the user's state.

The implementation of an adaptation engine in video games has normally been made using the player's performance as the input and a dynamic difficulty adjustment (DDA) as the output. This DDA consists of changing certain game parameters in real-time to adjust the game's difficulty level in order to keep the player optimally engaged, avoiding boredom (if the game is too easy) or frustration (if the game is too difficult). This optimal state, normally referred to as 'flow' [45], is the mental condition in which a person is fully immersed and engaged in a given activity. However, the adaptation engine has often ignored the role of affective states when interacting with a computer. This is especially relevant in gaming as video games often involve emotional processes that can have important effects on the player experience [24]. Adding affective metrics to the adaptation engine can substantially improve the effectiveness of the adaptation, providing a more personalised player experience.

Difficulty adaptation has been commonly used in games for entertainment [24], such as Yannakakis' *Bug Smasher* [172], which adapted certain game parameters depending on the player's interaction and performance in the game. Nevertheless, DDA has been used in the last few years for other purposes such as cognitive training [99]. *NeuroRacer* [7] was the first game-based cognitive training to include a performance-based adaptation engine that successfully improved the players' cognitive skills. The game analyses the player's performance and provides appropriate challenges depending on the player's progress. This shows that adaptation can have positive effects not only on the player experience, keeping players in a flow state, but also on the cognitive performance.

Other researchers have proposed the inclusion of affective metrics in the adaptation engine to improve the training efficiency [100]. Since the experienced affective states have significant effects on cognitive skills [19, 30], it is important to consider them when adapting a game for cognitive training. Using affective metrics, the adaptation engine would be able to provide a more personalised cognitive training. Although the advancements in wearable and motion sensors can provide an easy and reliable way of measuring physiological and behavioural signals, detecting and including these affective metrics in the adaptation engine is still a challenging task. The motivation behind this research is to improve the adaptation of game-based cognitive training programs by introducing affective metrics.

The recent introduction of affordable Head-Mounted Displays (HMDs) has increased research interest in VR. Although VR has been used in many different research fields [33, 163], this research is focused on its application to gaming and cognitive research, experienced through HMDs in contrast to PC screens or Cave Automatic Virtual Environment (CAVE). The use of HMDs for VR implies an embodied interaction that is crucial to achieve a sense of presence [72]. This sense of presence, defined as the feeling of '*being there*' in the virtual environment [131], involves high levels of immersion as well as an intensification of the experienced affective states [132, 131]. These high levels of immersion have a beneficial impact in cognitive skills [135] as they increase the degree of involvement with the given task [31]. The review by Rosa et al [135] of various VR-based cognitive interventions concluded that VR serious games have stronger effects than non-immersive interventions. They distinguished between three types of cognitive interventions: cognitive rehabilitation, cognitive stimulation and cognitive training. Whilst the first two interventions focus on "*the recovery of lost/altered cognitive functions*", cognitive training aims to improve the cognitive skills in healthy subjects. Following these findings, this research explores the benefits of high levels of immersion on cognitive skills in HMD-based VR games for cognitive training. In addition, since arousal and valence can have important effects on cognitive skills [19, 30], this work will investigate how the intensification of affective states reported in VR affects the player's cognitive skills.

According to the challenges and gaps in the related work presented, this thesis will address the following research questions:

1. *Which sensors are most suitable and reliable to measure the player's affective states in gesture-based gaming?*

2. *What is the impact of VR game playing on the player's affective states and WM performance?*

3. *What are the effects of arousal and valence on WM performance when playing a VR game?*

4. *Can adaptation improve the WM of subjects playing a VR game for cognitive training?*

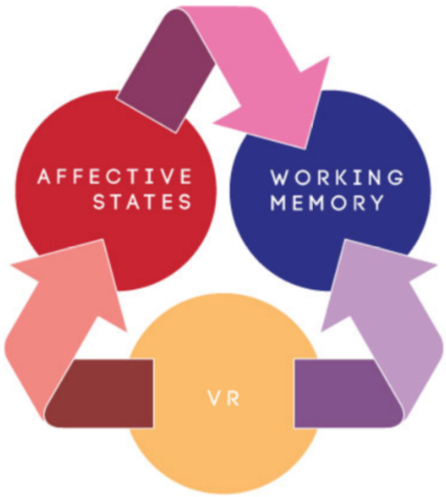

Figure 1.1: The three core factors studied in this research and the links explored between them

1.2 Aims and Approach

This research studies the links between three core factors: affective states, WM and VR. As shown in Figure 1.1, the effects of VR on the user's affective states and WM will be explored, as well as the influence of the experienced affective states on the user's WM performance. In order to address the above research questions, the work presented in this thesis has three principal aims:

1. *Investigate how VR gaming impacts on the player's WM performance and affective states compared to less immersive interaction modes such as traditional desktop computer gaming.* Previous research on the effects of VR on the user's affective states has demonstrated that the high levels of immersion often reported can intensify the experienced affective states [132]. Similarly, these high levels of immersion reported in VR can also have positive effects on the cognitive skills [135]. However, to the researcher's knowledge, no previous research has explored the effects of VR gaming on WM.

2. *Study the effects of arousal and valence on player's WM performance in a VR gaming setting.* Since affective states can have important effects on player's cognitive skills [19, 30], their WM performance could be enhanced by controlling for changes in the affective states or directly manipulating

them to achieve an optimal state. According to the Theory of Flow [45], a positive affective state can help subjects' cognitive performance. Although the link between emotions and cognition has been thoroughly studied in many different settings, mostly using images for emotion induction, it has not been studied in the fields of gaming or VR. Due to the aforementioned intensification of affective states in VR, caused by high levels of immersion [132], this research explores the effects of arousal and valence on player's WM in a VR environment.

3. *Explore the effects of DDA on player's WM performance in VR gaming.* Dynamic adaptation aims to sustain the players in an optimal affective state that would improve their WM performance [45, 19]. Adaptation methods used in video games have focused on performance-based adaptation, which evaluates the player's performance to make adaptation decisions. For example, the game *NeuroRacer* [7] has demonstrated that this approach can improve player's attention and memory on older adults, even six months after training with the game. This could be substantially improved incorporating the players' affective state in the adaptation loop [100]. Since arousal and valence have significant effects on cognitive skills, the introduction of affective metrics in the adaptation engine can provide a more effective and personalised adaptation and, therefore, better cognitive training. This is especially important in game-based cognitive training programs since video games often involve emotional processes [24] that can have significant effects on the player's cognitive performance. Although there has been little research on affect and performance-based adaptation for cognitive training [175], this thesis will explore its effects in a VR gaming setting in a longitudinal study.

1.3 Contributions

Following the aims described in the previous section, this work contributes towards the improvement of game-based cognitive training programs. The significant effects of affective states on WM performance highlight the importance of including affect metrics in the adaptation of games for WM training. This PhD thesis makes the following novel contributions to the existing literature:

1. *Game-based cognitive training should make use of VR to achieve maximal benefit.* This research further confirms that VR increases the self-reported levels of immersion and intensifies the players' affective states. These high levels of immersion are beneficial for player's WM performance, as it makes him or her more focused on the given task.

2. *Affective states can have positive effects in player's WM performance.* Although performance gains depend on the levels of arousal and valence, players in an optimal affective state, where both arousal and valence are high at the same level, can improve their WM performance.

3. *A novel methodology for affect recognition in VR gaming is presented.* The multimodal affect recognition system measures the player's head motion, heart rate (HR) and pressure exerted on the hands (muscle activation) to infer his or her arousal and valence levels. Although physiological signals have been previously used for affect recognition, this thesis presents a new methodology which, rather than adding new sensors for motion detection, employs the existing ones in a Head-Mounted Display (HMD) to measure the player's arousal and valence while playing in VR. Additionally, this methodology proposes the use of an electromyograph (EMG) to infer the player's valence, analysing the pressure exerted on the hand in gesture-based video games.

4. *A novel adaptation engine that combines affective and performance metrics for cognitive training.* Using the affect recognition system described above, this adaptation engine adjusts the difficulty of the game in real-time to provide personalised cognitive training. By combining performance and affective metrics to inform the adaptation engine about the player's progress and affective states, the adaptation decisions can be more accurate and effective. To the researcher's knowledge, this is the first game for cognitive training in VR using performance and affective metrics for dynamic difficulty adaptation (DDA).

5. *The use of game adaptation for cognitive training in VR is more beneficial for subjects with low WM capacity.* This finding is likely because affective states seem to have stronger effects on these subjects. This can be explained by the poor self-control of emotional experiences of these subjects [144], which makes them experience the negative effects of high levels of arousal (when not accompanied by high levels of positive valence) earlier on their cognitive performance. This contribution supports the importance of using affect and performance-based adaptation in games for cognitive training.

1.4 Associated publications

Parts of the work presented in this thesis have been published in international conferences, as follows:

- Chapter 3: Gabana, D., Tokarchuk, L., & Gunes, H. (2016). *Measuring affective, physiological and behavioural differences in solo, competitive and collaborative games.* In International Conference on Intelligent Technologies for Interactive Entertainment (pp. 184-193). Springer, Cham.

- Chapter 4: Gabana, D., Tokarchuk, L., Hannon, E., & Gunes, H. (2017). *Effects of valence and arousal on working memory performance in virtual reality gaming.* In Affective Computing and Intelligent Interaction (ACII), 2017 Seventh International Conference on (pp. 36-41). IEEE.

- Chapter 5 and 6: Gabana, D., Tokarchuk, L., Hannon, E., & Gunes, H. (submitted September 2018). *Affect and Performance-Based Difficulty Adaptation in Real-Time for Working Memory Training in Virtual Reality Gaming.* IEEE Transactions in Affective Computing.

1.5 Thesis outline

This thesis is structured in seven chapters, as follows:

Chapter 2 outlines previous research done in the areas of interest of this thesis - affective gaming, VR and game-based cognitive training - reviewing the intersections between these research fields. The chapter concludes with recent advancements in the adaptation on games for cognitive training.

Chapter 3 describes an initial pilot study of this research, where two co-located subjects played a Wii video game in three play modes (competitively, collaboratively and solo). Appropriate physiological and behavioural sensors for affect detection were explored in this study.

Chapter 4 presents a study that investigates the effects of game playing in two interaction modes (VR and Desktop) on players' affective states and WM performance. This study presents a first version of the custom game *Memory Break*, developed for this study for VR and Desktop settings.

Chapter 5 documents the design and development of the affect detection system using machine learning and the data collected in Chapter 4's study.

Chapter 6 reports the last study of this thesis, where the affect detection system and the adaptation engine implemented in *Memory Break* are tested in a longitudinal study with adaptive and non-adaptive versions of the game *Memory Break*.

Chapter 7 concludes this thesis summarising the main findings, contributions and limitations of this research. Future research directions are also proposed.

Chapter 2

Related Work

The topic of this thesis lies across a number of disciplines. This chapter discusses relevant work in the areas of interest, which can be categorised in four topics: i) affective gaming (section 2.1), ii) the effects of affective states on cognitive skills (section 2.2), iii) game-based cognitive training - concretely working memory (WM) training - (section 2.3), and iv) adaptation in games for cognitive training (section 2.4). These research areas and the intersections between them are discussed in detail.

2.1 Affective gaming

Affective computing is the research area started around 1995 in the MIT Media Lab. by Rosalind Picard. She defined it as "*computing that relates to, arises from or deliberately influences emotions*" [121]. In other words, it tries to give computers the ability to recognise and express emotions. This concept of making machines aware of the user's emotions has been applied to many different disciplines like robotics, healthcare, education or human-computer interaction [156]. [171]. This thesis is focused on its application to video games, known as affective gaming, for cognitive training.

2.1.1 Measuring affective states in video games.

The analysis of emotions in affective computing applications is normally explored in terms of affective states [62], which can be defined as the internal states experienced when feeling an emotion like happiness, sadness or frustration. These affective states unfold in two dimensions: arousal and valence (see Fig. 2.1) [138]. Arousal refers to the activation (or excitation) level of an individual, which ranges from high to low. Valence refers to the hedonic or pleasant

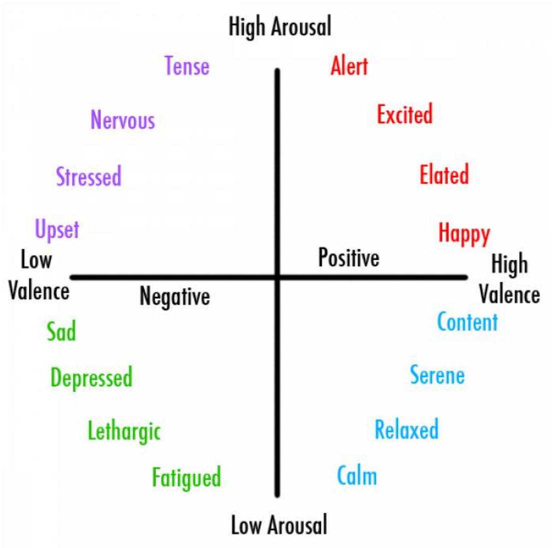

Figure 2.1: A circumplex model of affect [138]

(positive or negative) degree of an emotion.

One of the biggest challenges of affective gaming is how to detect and analyse the player's affective states when interacting with the machine. Öhman [111] distinguish between three modes to measure affective states:

1. **Self-reports:** user's subjective evaluation of his or her own emotion, frequently reported after the event through questionnaires.

2. **Overt behaviours:** Observable behaviours conveyed through gestures, body positions (postures), facial expressions, etc.

3. **Physiological responses:** Automatic reactions to a stimulus that triggers physical responses. There are many physiological indicators that can help to understand the underlying affective state of the user:

   - Electrocardiogram (ECG): relates to heart activity and is associated with emotional arousal, cognitive efforts and stress [158]. The most common features are heart rate (HR) and heart rate variability (HRV).
   - Electrodermal activity (EDA): also known as Galvanic Skin Response (GSR), measures the electrical conductivity through the skin. EDA is related to the sympathetic nervous system and is a good indicator of stress and anxiety [41].
   - Electromyograph (EMG): measures the electrical activity of muscle movements. For example, facial muscle activity is frequently used as

an indicator of valence [158, 139, 106].

- Brain activity: The measurement of neurological activity can be done using different technologies such as electroencephalography (EEG) or functional magnetic resonance imaging (fMRI) [24]. Although in its infancy, some studies have demonstrated the recognition of emotions and mental efforts through brain signals [106].

The detection of the affective states can be sometimes invasive to the players, limiting their movements or affecting their immersion level. Physiological sensors have to be physically attached to the player's bodies to measure the physiological changes evoked by the affective states experienced. However, the development of new technologies and novel approaches in the past few years allows a less invasive measurement of the player's affective states. Recent advances in wearable technology have successfully built portable devices that remove the sensation of having a sensor attached, giving the player greater freedom of movement [81]. The progress of Computer Vision techniques to identify behavioural cues such as body movements, gestures or facial expressions [13], enables affective computing to recognise user's emotions in an easier and more reliable way. For example, Savva et al [143] used a motion capture (see Fig. 2.2) system to analyse the affective states of the users playing a Wii tennis game. Using a machine learning algorithm to automatically recognise the gestures, they grouped the emotions conveyed into four categories: '*high intensity negative emotion*', '*happiness*', '*concentration*' and '*low intensity negative emotion*'. The results were compared with the classification of two independent groups of observers, showing the system's accuracy to be 57.4%. On the other hand, Mandryk and

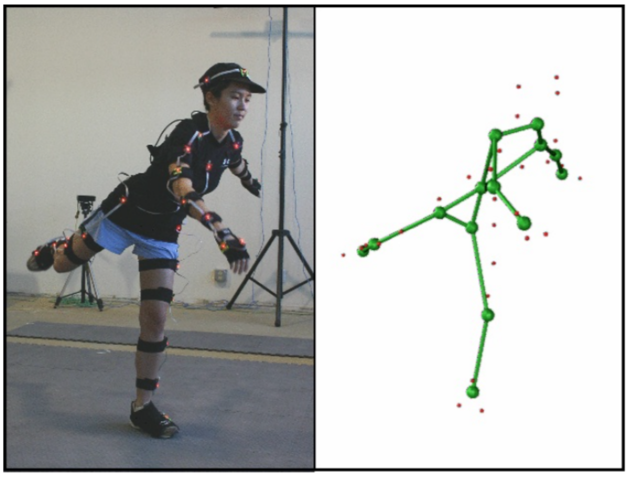

Figure 2.2: Motion capture system to analyse behavioural cues.

Atkins [89] used physiological signals (facial EMG, ECG and skin conductance) to measure participants' affective states while playing a video game alone and against a friend. Using a fuzzy logic model, they transformed the physiological responses into continuous arousal and valence levels. They found significant correlations between the detected affective states and subjective self-reports.

It has been demonstrated that body movements in games depend significantly on the player's motivations [108]. When the main motivation is winning, players use full body movements to exploit all the functionalities of the controller. This means that by recognising body movements in games we understand the quality of the user experience [89]. Due to individual dissimilarities on physiological and behavioural signals, it is important to normalise the data obtained [139, 158]. Savva et al [89] normalised participant's behavioural data while playing a Wii tennis game using the maximum and minimum values of each body part over the whole data set. Physiological signals should be normalised using the individual baseline of each participant, usually obtained by asking the participant to relax for a few minutes. Thus, the normalized data would show the amount by which a participant increased his or her physiological responses in each experimental condition.

Affective gaming has traditionally used the detected affective states to improve the player experience by either controlling the interaction with the physiological signals manifested by the player (biofeedback) or adapting the game according to its objectives and the player's current affective states (adaptive feedback) [139, 58]. Bersak et al [20] developed 'Relax-to-Win', a biofeedback racing game where two players compete for control of a virtual dragon. Each player wears a GSR sensor that measures the stress level. When the player relaxes, the skin conductance decreases, making the dragon fly faster. Alternatively, if the player gets stressed the dragon slows down its pace. Chanel at al [37] made an adaptive version of the popular game Tetris which automatically changed its difficulty level depending on the emotions recognised (boredom, anxiety and engagement). It did this through the assessment of 5 physiological readings: GSR, ECG, EEG, respiration rate and skin temperature. The authors proposed two aims for emotion assessment in games: 1) evaluate the game from a user-centred perspective, and 2) maintain player's engagement through difficulty or content adaptation to induce particular emotions.

2.1.2 VR as an affective medium

The effects of video game playing in engagement and immersion has been widely investigated. Many researchers have studied these effects in different contexts such as learning [140], dynamic adaptation [106, 37] or its influence in player's affective states [106, 95, 16]. Although these two concepts are related, not many authors have studied the distinctions between them. Brown and Cairns [31] interviewed some gamers about their experiences when playing video games. They defined immersion as "*the degree of involvement*" with a computer game and identified three levels of immersion:

1. **Engagement:** is the lower level of immersion and must occur before the rest of the levels. Players have to put effort, time and attention to engage with a game. Engagement is sometimes referred to by players as being focused.

2. **Engrossment:** as the immersion increases, players become less aware of their environment. This immersion level is achieved when the game features directly affect the player's emotions.

3. **Total immersion:** often called presence, the highest level of immersion is described as a complete detachment from reality, where players are entirely absorbed in the game, forgetting everything around them.

The concepts of immersion and presence has been often used synonymously in different contexts [95]. Nevertheless, the term presence has been frequently used in Virtual Reality (VR) environments, especially in video games, to define the feeling of '*being there*' in the virtual world [131]. The feeling of presence elicits such a realistic interaction that users feel they are inhabiting the virtual environment [164] to the point that virtual experiences can evoke similar perceptual reactions and emotions as in the real world [85, 137]. Thus, VR has sometimes been called an '*experiential interface*' [164]. The interaction between presence and emotions in VR has been examined in depth in a number of studies [16, 164, 132], all of them concluding that a higher level of presence influences directly the vividness and intensity of the emotions experienced. Sometimes these studies have also referred to VR as an '*affective medium*' [164, 132], due to its possibilities to evoke target affective states.

Many researchers claiming to use VR in their studies have employed different platforms to deliver VR content. Three main platforms are normally used: PC screens or monitors, head-mounted displays (HMD), and immersive virtual environments where the user can walk around a room with projections on the walls, floor and ceiling (i.e.: Cave Automatic Virtual Environment (CAVE)). These three platforms to experience VR are very different and induce distinct affective reactions and presence levels on the users. Several researchers have studied the affective states evoked by VR, mostly experienced through PC screens [167, 22, 18] or comparing two or more of the VR platforms aforementioned. The reason to use PC screens is normally to prevent motion sickness in their participants, often people with some kind of mental health diagnosis such as autism [167, 22]. The recent introduction of affordable HMDs and the improvements in this technology over the last few years has considerably reduced the risk of motion sickness, which may encourage the use of these devices in VR research. This thesis therefore will focus exclusively on the use of HMDs for VR experiences.

Kim et al [74] explored participants' emotional responses to these platforms for VR content delivery. They found that while PC screens evoke low arousal, HMDs and CAVEs can produce higher or similar levels of arousal. Regarding the valence dimension, HMDs and CAVEs evoke different levels of valence, being positive for CAVE and negative for HMD and sometimes for PC, depending on how stressful the virtual environment is. The reason why valence was negative in HMD in this particular study by Kim et al [74] was because the device caused motion sickness, probably due to the technology not being fully developed at that time (2014). In regard to immersion, CAVEs and HMDs clearly provoked higher levels of presence than PC screens. Riva et al [132] explored the effects of mood induction in presence creating three virtual parks associated with three emotions: anxious, relaxing and neutral. Participants self-reported their emotions and interacted with the VR environment using a HMD and a controller. Their results demonstrated the reciprocal interaction between presence and emotions: whilst the feeling of presence was higher in the emotional environments (anxious and relaxing), the emotional state also influenced the presence levels. Other research has looked at the physiological responses wearing a HMD. Meehan et al [96] tested the physiological changes using passive haptics (i.e. a wooden ledge) to explore its effects in player's presence levels. They created a virtual scenario with 2 rooms (Fig. 2.3), one of them with a pit (Pit Room), where the user, with a controller and wearing a HMD, had to move objects to the Pit Room standing on a wooden ledge (Fig. 2.3). Results confirmed that the presence of passive haptics significantly increased participants' heart rate (HR) and level of presence.

Besides the studies using physiological signals to measure affective states, there is very little research that uses gestures or motion to detect player's affective states in VR wearing a HMD. Becker-Asano et al [17] analysed participants' head movements in a HMD-based VR emergency situation to infer the emotional arousal. Based on previous research about affect recognition of head movements [54], they assessed the head's pitch and yaw rotation as well as the interaction with the joystick controller, successfully mapping the horizontal movement speed to the player's arousal level measured through physiological signals. Beyond this paper, it seems there is not much academic research about inferring affective states looking at head gestures in VR. However, very recently,

Figure 2.3: View of the virtual environment (left) used in the study by Meehan et al [96] and a participant of this study moving an object to the Pit Room standing on the wooden ledge (right) [96].

many companies especially in Sillicon Valley [1], have shown interest in emotion detection using HMD VR. Retinad Analysis [2] specializes in measuring emotional reactions to 360 videos in VR, trying to understand how the human head moves when feeling a certain emotion. Other companies employ different technologies, for example, Yotta Technologies tracks facial muscles and eye movements to capture microexpressions and more subtle emotions, or the Korean Binary VR that uses cameras inside headsets for facial recognition. The British company EmTeq [3] wants to use affective computing for healthcare and entertainment in VR using biometric sensors that would be installed in the HMD's faceplate and eye-trackers inside the device to accurately detect subtle facial movements. Consequently, emotion detection can be used to understand affective responses to improve behavioural and mood self-control.

2.2 Affect and Cognition

According to the definition of the Oxford dictionary, cognition is the "*the mental action or process of acquiring knowledge and understanding through thought, experience, and the senses*". It includes multiple cognitive processes like attention, memory, reasoning, evaluation problem solving or decision-making [99]. Attention is defined as the faculty used to regulate the flow of information into

[1]http://fusion.net/story/312561/virtual-reality-emotion-surveillance/
[2]https://www.retinadvr.com/
[3]http://emteq.net/

the sensory system, prioritizing the relevant components and eliminating the distracting inputs [157]. There are different types of memory, all of them with a limited capacity. While short-term memory refers to the amount of information a person can retain for a brief interval of time, long-term memory refers to the retention of information for longer intervals [147]. In addition, working memory (WM), defined as the cognitive function that temporarily stores information while other complex cognitive activities are in course [64], holds and manipulate novel information, being an important executive function for reasoning and decision making [49].

These cognitive functions and emotions have a bidirectional relationship. Emotions can determine how we perceive the world guiding our attention, organising our memory or influencing our decisions [30]. Emotional stimuli can facilitate perception, drawing the attention more quickly and impeding disengagement for a longer period than neutral stimuli [30]. Additionally, cognition influences emotions by allocating cognitive resources for information processing and defining goals [164]. Some studies claim that relevant information for our well-being have a preferential encoding in memory in order to influence our future behaviour [102] and emotionally charged simtuli at risk of being neglected [157]. Thereby, stimuli with negative valence are faster to process as they may have an imminent threat to survival [53]. Bennion et al [19] summarised the three main hypotheses about how emotion affects memory:

1. Emotion usually enhances memory.
2. When it does not, is can be explained by the physiological arousal, benefiting memory to a point but then having a detrimental influence [174].
3. When emotion helps to process information (encoding), it also facilitates the storage of that particular information (consolidation).

The amygdala has an important role on the attention and codification of emotional stimuli. This part of the brain is responsible for dealing with emotional reactions and certain cognitive functions [6]. It is fundamental in encoding emotional information, being able to modulate visual processing and augmenting the possibilities to identify and attend emotional stimuli in the environment [71]. Neuroimaging studies (i.e. producing images of the brain activity) have observed that events with higher emotional charge evoke greater activity in the amygdala and that these events are more prone to be remembered [120].

Although certain research has claimed that the amygdala responds primarily, but not only, to the arousal and not the valence of information [14], it is true that positive and negative emotional stimuli are more likely to be recalled than neutral non-emotional stimuli [71]. In 1908, Yerkes and Dodson [174] found that arousal can have positive effects on memory up to a point, after which it has negative effects if too high. However, the effects of arousal may depend on whether the arousal experienced is relevant or related to the memory task. Libkuman et al [86] found that physiological arousal produced while running or cycling does not impact the memory of a scene's details. Consequently, the effects of arousal may depend not only on its intensity but also on other factors inherent to the given task.

Whilst the arousal dimension of affective states can be decisive for the memory enrichment of emotional stimuli, changes in the valence dimension, without altering arousal, can also improve memory vividness [71]. Positive and negative valence stimuli have different effects on memory encodings, though there is certain controversy about its specific influence. Studies using words or images to evoke an emotional reaction on the subject have concluded that negative items are more likely to be remembered than positive [113]. Conversely, studies evaluating self-reported recall of personal memories proved that subjects remember positive events better than negative ones [11]. Thus, individuals looking for positive states have better memory for positive than negative stimuli. Other authors like Storbeck and Clore [152] have suggested a more detailed explanation of the effects of valence on memory. They suggest that negative affect leads to a more focused and deeper processing of details of information, and positive affect brings a broader and wider attention. In summary, whilst it has been proved that arousal affects memory, enhancing the encoding and vividness of memories, the role of valence is controversial and still not clear.

Even though further research needs to be done in this area, some researchers have come up with different ideas for its practical applications. For example, Wadlinger and Isaacowitz [165] proposed that we could train our attention to regulate emotions, using meditative practices to train our selective attention to focus on positive information and enhance our well-being. Moore et al [104] reviewed the effects of emotions and motivation on WM during math problem solving. Some studies examined used a similar approach to the AOST (see section 4.2.1), mixing simple math problems and letters that have to be remembered as a secondary task [146]. This review concluded that people showing higher levels of anxiety solving math problems performed considerably worse than those with lower anxiety. Thus, mathematics anxiety occupy WM resources required to solve problems successfully [104]. Ashby and colleagues [12] reviewed the influence of positive affect in creative problem solving. They assert that "*there is substantial reason to believe that the effects of positive affect and arousal are not identical, and that the well-documented improvements in creative solving problem that occur under positive affect conditions are indeed due to induced positive affect, and not simply to increases in arousal*". Yeh et al

[173], investigated how negative emotions influence WM on a creativity game. Similar to Bennion et al [19], their findings indicated that arousal could help WM to a point but then exert negative effects. For this reason, they propose a game that presents challenges and stressors that would activate the player's attention to an optimal level whilst avoiding provocation of negative emotions [173].

Whereas arousal and valence can have distinct effects depending on the task and context, subjective factors may also influence the impact of emotions in our memory. Curci et al [46] studied how individual differences in WM capacity (i.e.: high versus low capacity) and the emotional valence of the stimuli influence WM performance. They presented negative and neutral emotional stimuli before and after performing the Operation Span Task. Their results indicated that participants with low WM capacity had a worse performance than those with high WM capacity when presented negative valence stimuli. Therefore, the emotional valence of stimuli and the subject's WM capacity affects the WM performance. Similar results were found by Schmeichel and his colleagues [144] in a series of four studies. They concluded that subjects with high WM capacity suppress negative and positive emotions better than those with low WM capacity when watching movies. These findings indicate that the effects of arousal and valence on WM not only depend on the task's nature but also on individual factors.

2.3 Game-based cognitive training

Video games have been used for many other purposes than just entertainment, such as education or cognitive training. They have been proved to be a good conductor to keep the player engaged and focused on a task [93]. Most of today's commercial video games are made of carefully designed storylines, artworks and soundtracks whose primary goal is to completely immerse and engage the player [8, 98].

Cognitive training exercises often lack engaging and fun elements that keep the user motivated. Using gaming elements such as immediate feedback on performance or a story line has been demonstrated to improve not only user's motivation but also performance and transfer effects [8]. Prins et al [124] studied the effects of intrinsic and extrinsic motivation in cognitive training, giving different monetary incentives (extrinsic motivation) and using a gamified version of the WM exercise given (intrinsic motivation). Results showed that the group with the strongest incentives (10 euros and gaming version) outperformed those in the low-motivation group (no incentives). Many researchers have suggested that intrinsic motivation plays an important role on the success of game-based learning and cognitive training programs. Intrinsic motivation can be defined as the inherent motivational value of a content or task that does not require external incentives [130].

Lepper and Malone [88] investigated which elements, particularly in computer games, make learning fun and intrinsically motivating. They studied what makes children prefer certain games, concluding that it depends on whether the game has 1) explicit goals, 2) a score, 3) audio effects, 4) random elements and 5) if response speed made a difference in the game. They proposed a taxonomy of three characteristics that make activities intrinsically motivating:

- **Challenge:** Activity should provide goals and continuous performance feedback to increase the player's motivation and self-esteem. Proximal goals can create higher motivation and better task performance than distant goals.
- **Curiosity:** Is the desire to learn or explore new things and the most related to intrinsic motivation. Malone and Lepper distinguished between sensory (perception of stimuli) and cognitive (higher-level mental structures) curiosity.
- **Fantasy:** Refers to elements that make a meaningful context and let players augment their imagination.

Challenge and curiosity are closely related and both need to be optimally sustained [130]. Games should provide adequate levels of challenge that stimulate player's curiosity, increasing the motivation and task performance. Challenge levels can be changed manipulating the difficulty level or the goals. Therefore, it is important to have an appropriate sequence of difficulty levels that progress as the player's skills and competence increase [88]. Even though the aesthetics and story line of a game are interesting for some players, the optimal difficulty level to keep them challenged at the right level can change considerably between players as it depends on their skills. This is closely related to Csikszentmihalyi's Theory of Flow [45], which argues that challenge should be adjusted to the player's skills in order to provide an optimal experience (Figure 2.4). Thus, the difficulty adjustment in game-based cognitive training programs is crucial for its success. Keeping players challenged at the right level leads to an optimal experience that can have positive effects in players' performance [8].

Klingberg [77] suggested three factors that contribute to a successful cognitive training program. First, training should not train specific strategies for just remembering more information. Second, training should be specifically focused on WM tasks. Finally, training should adapt to the user performance. The target of cognitive training is often an improvement of fluid intelligence, also known as transfer intelligence, defined as the ability to solve new problems and reason abstractly [68]. In other words, it is the ability to apply the skills acquired during training to other domains or situations [147]. There are two types of transfers. While near transfer refers to a performance improvement on tasks that measures the same skill trained, far transfer refers to improvement on abilities related to those trained [147]. It is important to distinguish between these types of transfer and the practice effects, described as "*improvements in cognitive test performance due to repeated exposure to the test materials*" [51].

2.3.1 Not all games are created equal

The employment of video games in the research of cognitive training can be divided in two categories: off-the-shelf games and gamification of existing cognitive tasks [98]. Some researchers have proven that certain commercial action video games can improve specific cognitive skills like visual attention [110, 43]. Cohen, Green and Bevalier [43] have studied the effects of different video games in visual attention. Participants (18 to 29 years old) with little to no gaming experience were trained on one of several video games (*Unreal Tournament 2004*, *America's Army*, *Harry Potter* or *Tetris*) for 12 hours and tested on two paradigms: attentional blink and multiple object tracking. Individuals who played the action video game (*Unreal Tournament*) showed improvements on these two measures of visual attention while participants trained on other games showed less or no improvement. They concluded that fast paced games and requirements to track various objects at once, produce changes in visual selective attention. Although participants trained in other games showed significant im-

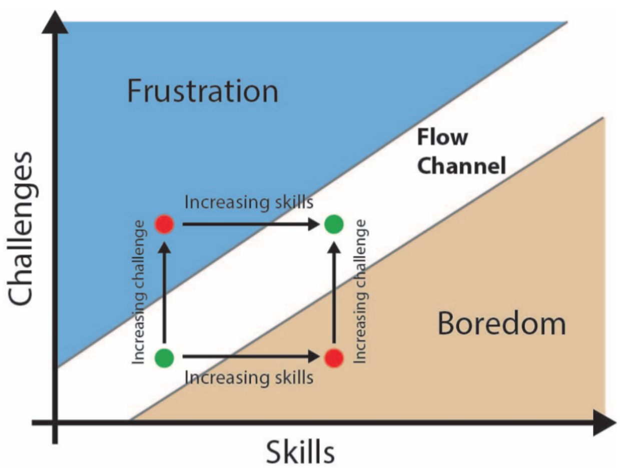

Figure 2.4: Flow channel according to Csikszentmihalyi [45].

provements on the visual attention tests, they could not conclude that it was produced by the training itself as they could not compare it with a control group that was not trained on any of the games. Oei and Patterson [110] also tested the benefits of four different commercial action games on attention and WM played on iPhone/iPod. Their study demonstrated that not all games have the same cognitive demands, and transfer to a cognitive skill is more likely if the underlying skill is highly practiced in the game [8, 110, 77].

On the other hand, the gamification approach simply introduces game elements like storyline, goals or rewards in boring cognitive tasks, making slightly more engaging exercises [98]. This is a very successful method in certain populations like children [160]. Anguera and Gazzaley [8] argue that maybe the low-engaging nature of cognitive tasks has driven the usual negative results in the field of game-based cognitive training. They explain that one of the most important factors differentiating these two types of games is the motivational factor. Most commercial video games create a high engagement and immersion through carefully designed stories and reward structures that challenge and motivate the individuals to reach greater outcomes. Due to their targeted approach, these cognitive exercises are more focused on challenging specific cognitive skills than including engaging game elements to motivate the player [8]. The introduction of game elements to increase engagement and intrinsic motivation can benefit the players' attention performance, creating a more effective cognitive training [82]. These elements should be carefully incorporated as they can also bring distractions, even though some research suggests that off-task behaviours are a way for players to regulate their emotions [140].

This gamification method was very popular a few years ago when private companies such as CogMed or Lumosity, developed specific training programs based on the gamified versions of existing verbal and visuo-spatial cognitive tasks [147]. Although these programs have been tested in different studies with positive results [92, 64, 78], many researchers have constantly questioned the scientific validity of these experiments [147] demonstrating that these brain training programs do not improve our cognitive skills [115]. A consensus signed by many well-known researchers on the so called Brain Training industry published in October 2014 [4], express a general agreement on the fact "*there is no compelling scientific evidence to date*" that *brain games* reduce or reverse the cognitive decline in older age population and patients with Alzheimer's disease. In January 2016, the company Lumosity has been fined by the American Federal Trade Commission to pay 2 million dollars for deceptive advertising. Lumosity claimed that its online mini-games can improve cognitive skills in older population and even ward off Alzheimer's disease [4]. It is important to note that these

[4]https://www.ftc.gov/news-events/press-releases/2016/01/lumosity-pay-2-million-settle-

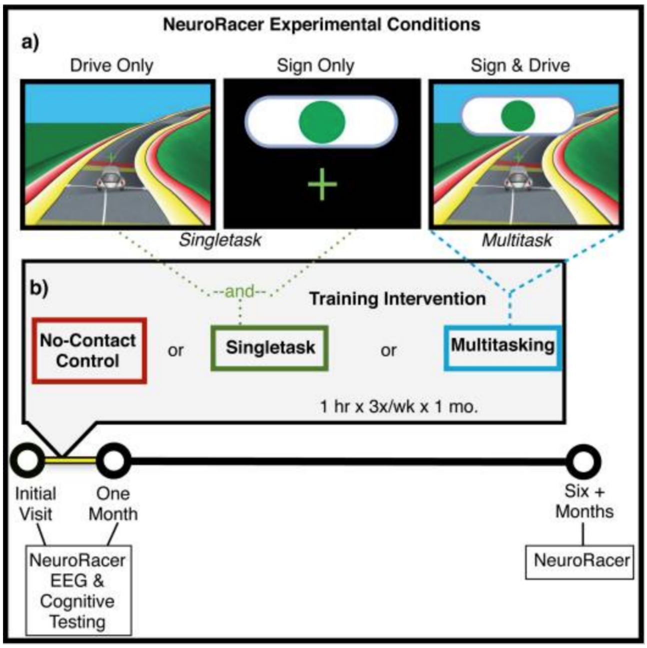

Figure 2.5: *NeuroRacer* experimental conditions (A) and training design (B) [7].

two events are focused on the use of *brain games* to cure Alzheimer's disease and cognitive decline in older people. Further research needs to be done to clarify whether video game playing can improve our cognitive skills.

Gazzaley and his research group have created a game called *NeuroRacer* [7] that claims to enhance multitasking in older healthy adults. In this game the players have to drive a car while occasionally responding to the coloured shapes presented (Figure 2.5). Playing an adaptive version of *NeuroRacer* in multitasking mode for one month, adults between 60 and 85 years old reduced the cognitive costs when task-switching in multitasking mode and improved the sustained and divided attention as well as their WM. These effects persisted 6 months later. Even more revealing is the recent work by Shute, Ventura and Ke [148], which claims that playing the puzzle video game *Portal 2* for 8 hours, led to more positive effects in spatial and other non-cognitive skills than training the same time on multiple mini-games on the Lumosity platform. This mean that it may be dependent on the game's design and focus, whether it is possible to improve cognitive skills like attention or WM. Anguera and Gazzaley [8] argue that the higher cognitive impact of video games compared to gamified cognitive exercises might be due to their fun and motivational factor.

Even though further research is required to demonstrate that game-based

ftc-deceptive-advertising-charges

training can improve with cognitive skills [147], it is vital to highlight some important facts. Firstly, the impact of these cognitive games in different groups of people like novice and expert video game players must be clarified. Some researchers have argued that skilled players have better visual attention skills [59] and their extent of improvement in video game training is smaller compared to novices [141]. Hence, greater effects from video game cognitive training would be observed in novice players than experienced ones [59, 141]. Even more important is the age of the subjects. Toril et al [159] completed a meta-analytic study to investigate the effects of video game training in older healthy adults compared to younger adults. They concluded that video game cognitive training has positive effects in older adults, with age and number of sessions being significant in modifying the effect size of the training program. Finally, as aforementioned, certain personal factors like motivation to play or level of fun experienced are modifiable variables that can explain the inconsistency of results [159, 8].

In summary, it is not clear whether game-based brain training programs can improve cognitive abilities in all populations. Recent studies have rejected this hypothesis [115, 4], especially when using programs like Lumosity, which claim that playing 15 minutes 3 days a week we can improve our attention or WM. However, it seems more evident that certain entertainment video games (especially action games) can improve certain aspects of our cognition like visuospatial attention [110, 43]. Results like the ones achieved by the game *Neuro-Racer* [7] or by Cohen, Green and Bavelier [43] encourage the idea that it depends on the game itself whether it can improve specific cognitive skills. On top of this, other studies point out the effect that subjective factors like motivation can modulate the effects of these games on our cognition [159, 8].

2.3.2 Cognitive training in VR.

Virtual Reality (VR) has been widely used in cognitive rehabilitation, assessment, stimulation and training [135, 114]. A review of 151 papers about serious games in VR for cognitive interventions by Rosa et al [135] claimed that VR has consistent positive effects in our cognition, especially in attention and memory. This review also concluded that VR-based serious games for cognitive interventions are more effective than traditional non-immersive approaches [135, 109]. When subjects feel physically present in a virtual environment, the transfer of cognitive and behavioural skills into the real world increase [125]. In addition, one of the reported advantages of VR applications for cognitive interventions is the possibility to replicate real world tasks in a controlled way, where all aspects can be manipulated and adapted to the user [135]. Chittaro et al [39] developed a HMD VR game for aviation safety education. They assessed participant's knowledge retention and physiological arousal in the game compared to a standard aviation safety educational method (the safety cards). Their results showed that the VR game produced a higher knowledge retention than the safety cards, as well as higher levels of engagement and arousal, two factors that could contribute to better knowledge retention.

It is important to note that not all the studies that claim to use VR or virtual environments in their studies have actually used immersive VR technologies such as HMDs, using instead PC screens or CAVEs to experience the virtual environments [114]. As reported earlier in this chapter (see section 2.1.2), HMDs and CAVEs produce higher levels of presence and arousal than PC screens. This important difference in the way the virtual environments are experienced should be considered when examining the effects of VR in cognitive training.

One of the issues that garners most attention from neuroscientists and psychiatrists in the use of VR is related to its ecological validity for the assessment of cognitive abilities [107]. Ecological validity, in this case, refers to whether one can generalise the observed behaviour shown in the virtual environment to the traditional real world assessments [145, 109]. Matheis et al [90] used HMD-based VR to evaluate the ecological validity of this medium to assess learning and memory in both healthy subjects and those with traumatic brain injuries (TBI). Their findings reported a significant correlation between subjects' performance in VR and the California Verbal Learning Test, a standard neuropsychological measure of memory. They confirmed the ecological validity of VR for measuring learning and memory as they could differentiate between healthy and TBI subjects by analysing subjects' performance. A similar study by Ouellet et al [114] evaluated the ecological validity to assess everyday memory in a Virtual Shop. Two groups of adults with and without cognitive decline had to memorise a shopping list of common items and obtain them from a virtual shop using a "*natural navigation mode*" (i.e.: walking around the shop). Correlations with performance of existing questionnaires assessing every day memory supported the ecological validity of the Virtual Shop to differentiate between the two groups looking at their performance.

Another interesting study was conducted by Rizzo et al [133] and later extended by Nolin et al [109]. The first study by Rizzo et al [133] assessed the attention of children with Attention Deficit and Hyperactivity Disorder (ADHD) in a virtual class using a HMD and tracking the non-dominant hand and opposite leg, although subjects could not see their body in the virtual environment. Participants, interacting with a physical mouse, had to perform a Continuous Performance Task (CPT) where a series of letters were presented in the classroom's blackboard (one at a time) and subjects were instructed to respond hitting the mouse only when the letter '*X*' was preceded by an '*A*'. Audio and visual stimuli typically present in classrooms (i.e., ambient noise, paper airplanes flying around the room, human avatars walking into the room, activity occurring outside the window) acted as distracters to participants in two of the three conditions assessed. Reaction time to hit the mouse was recorded to assess performance measures as well as the head, hand and leg movements to evaluate the hyperactivity often present in ADHD. In a second study conducted by Nolin et al [109] sixteen years later, the validity and reliability of the tool and the relationship between performance and presence were investigated, although this time hand and leg motion were not tracked.

These measures taken and the VR proved to be a good assessment tool for attention and ecologically valid. Head movements (left-right, up-down and tilt) during the virtual test presented good correlations with traditional CPT performances, accounting for 76.12% of its variance and, thus, a good indicator of the participant's distraction level. The performance scores only accounted for a 12.11% of the variance and the level of presence did not show any correlation. They pointed out the benefits of HMD VR for the study, assessment and possible rehabilitation of attention disorders and for being a controlled environment for the manipulation of complex test stimuli in cognitive exercises.

2.4 Adaptation in games for cognitive training

One of the main goals of video games is to create enjoyable experiences that keep players engaged and motivated to play [8]. Motivation and player engagement are two important factors for the success or failure of any video game for entertainment or cognitive training. When players are in a flow state, also known as optimal experience, they make better use of their cognitive resources, having a better cognitive performance [8]. This has been evidenced by significant positive correlations of motivation and engagement with cognitive performance [124, 159]. Traditional approaches to sustain players in this flow state have consisted of the dynamic adaptation of certain game elements or parameters such as the difficulty level. Keeping the right balance between difficulty and player's skills, it is possible to provide an engaging and positive experience that can have positive effects in player's cognitive performance [101]. Thus, it is important to provide positive reinforcement in video games to maintain high motivational and engagement levels [103].

2.4.1 Adaptation in games.

Commercial video games have been using dynamic adaptive mechanics for many years. Video game designers and developers have always tried to make more engaging games by changing certain game elements to challenge the player. This adaptive method, often called Dynamic Difficulty Adjustments (DDA) [87], tries to balance the game's complexity to keep the player in a flow state (Figure 2.4) [45] by tweaking certain game parameters such as the difficulty level in real-time depending on the user's performance. Some researchers have called this type of video games Closed-Loop Video Games (CLVGs), defined by Mishra et al [98] as "*interactive video games that incorporate rapid, real-time, performance-driven, adaptive game challenges and performance feedback*" (Figure 2.6). The adaptation loop is often used to create a more immersive and engaging player experience keeping a good balance between difficulty and skill levels. As players advance in the game, developing their skills and making good progress, the game dynamically increases the difficulty to keep players engaged and motivated to play. One example of DDA is the rubber-band effect [116] applied in many games, especially in racing games. This effect consists of decreasing the speed of the player's car while leading the race and increasing it when falling behind.

There are many game elements that can be dynamically adjusted, though they should be adapted in a subtle way to avoid the player realising the changes. Gilleade and Allanson [57] propose some game elements that can be manipulated when adapting the game:

- **Difficulty:** Traditionally, users set this up (easy, normal or hard levels). However, games can have different difficulty levels in various elements such as speed, health of enemies, player's power or frequency of rewards among others.
- **Story:** Controls the strength of the drama involved in the story line. Different stories can be told depending on certain factors such as interactions, items taken or player's emotions.

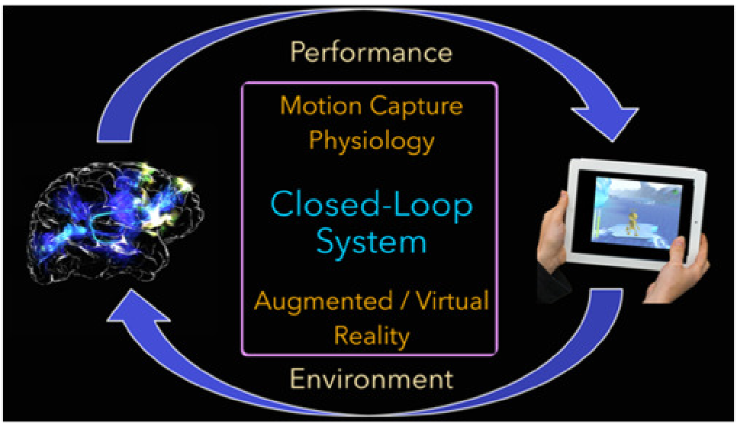

Figure 2.6: Diagram of a Closed-Loop Video Game (CLVG) [98]

- **Music:** Tempo (speed) and style.

CLVGs are still in its infancy, but its potential to include individual real-time measures to lead the game can increase player's engagement [98, 8]. Although CLVGs are normally based only on player's performance metrics, recent studies propose to include physiological or behavioural metrics that reinforce the adaptation. According to Mishra [98], the input channel of CLVGs should use not only the user's performance metrics but also real-time data of the player's interactions and behaviour. Using physiological or motion sensors to measure the player's HR, muscle activity or even eye movements can result on a more accurate adaptation loop that adjusts itself depending on a number of variables such as the player's behaviour, affective states or performance. Rather than a biofeedback approach, they envision a multimodal CLVG that integrates real-time physiological and behavioural data [98]. Since video games often involve emotional processes, it is important to consider the player's actual affective states to provide a more effective adaptation [24]

Liu et al [87] compared performance and affect-based DDA in a game designed to reduce anxiety, using multiple physiological signals (i.e., ECG, EMG, GSR, skin temperature, etc.) to detect players' anxiety levels. Three levels of performance (poor, good and excellent) and three levels of anxiety were identified to adjust the difficulty of the game in each DDA method. They found that affect-based DDA was perceived as more challenging but also more satisfying than the performance-based DDA. A similar study by Bontchev and Vassileva [25] also compared performance and affect-based adaptation in an educational video game. Using GSR and facial expressions for affect recognition, the adaptation improved from 64% when using only performance to 82% when combining both affect and performance-based adaptation. These findings suggest that affect-based adaptation can be as good as performance adaptation, though the combination of both can result in a more accurate adaptation [24]. Affect-based adaptation is usually implemented in video games using positive or negative feedback control mechanisms [24]. Whilst negative feedback control tries to reduce the distance between the player's current affective state and the desired emotional state, positive feedback aims to enlarge the difference between the detected affective state and the target state (i.e., trying to avoid frustration). Some games implement a hybrid approach, mixing positive and negative feedback to keep the player in a flow state.

One of the biggest challenges in affect-based CLVGs is how to recognise and measure player's affective states. As previously mentioned, researchers often use physiological and behavioural signals to detect changes in affective states. Multiple features are extracted from these signals to train a machine learning algorithm for affect classification or regression [38]. However, it is difficult to create subject-independent algorithms that are suitable to recognise the affective states of all subjects. Parsons and Reinebold [119] tested subject-independent and subject-dependent algorithms for arousal classification. While the subject-dependent algorithm was trained and tested using data from the same subject, the subject-independent was trained using data from other subjects. Results revealed that the subject-dependent algorithm achieved a classification accuracy of 96.5%, while subject-independent classification got an average accuracy of 36.9%, close to the chance level.

Although DDA has been extensively used in entertainment video games to sustain player's motivation and engagement, it has also been implemented in games for cognitive training. Since task difficulty directly affects cognitive performance [175], it is important to provide appropriate challenges to maximise player's performance. Hence, CLVGs can create personalised training programs that highly engage players in order to achieve cognitive improvements in the targeted skills and a transfer of benefit to other cognitive functions. Adding information about the player's affective states to the CLVG can bring a more accurate and personalised adaptation that would result in a better cognitive performance.

2.4.2 The next generation of games for cognitive training.

Previous studies where subjects repeated trials of one cognitive task without DDA only lead to faster reaction times and higher accuracy [79]. The repetition of one cognitive task just makes subjects better at that task, but no improvements on the trained or other cognitive skills are observed. Adjusting the difficulty of the task on a trial-by-trial basis according to the user's performance, and therefore pushing the limits of the cognitive skills, could enhance the trained cognitive skill [77]. For instance, Anguera et al [7] with their game *NeuroRacer* proved that only when the difficulty level is adapted to the player in the multitask version of the game, benefits appear to transfer to untrained cognitive skills in older adults. Mishra et al [100] also demonstrated an improvement in attention in older adults using an adaptive game that specifically challenges the player by increasing levels of distractions. Implementing DDA to constantly push a specific cognitive skill over a sustained period of time might induce the improvement of these abilities or even transfer to other untrained skills [7, 64].

Although DDA is normally implemented based only on player's performance metrics, CLVGs should include multimodal information such as physiological signals that supply the adaptation a good understanding of the player's cognitive and affective state [98]. Zhang et al [175] looked at the cognitive load of a VR driving task in teenagers with autism. They tested various machine learning algorithms in the classification and fusion of different features extracted from physiological, EEG and eye gaze as well as performance measures. The K-Nearest Neighbours (KNN) algorithm yielded the best results and was proposed, but not used, to adapt the game difficulty according to the player's affective states to provide fruitful skill development. The manipulation of the task's difficulty was made by increasing the number of directions and obstacles in the driving task. The authors found that the different difficulty levels of the driving task correlated significantly with the arousal level determined through physiological signals. Thus, they concluded that higher difficulty levels increases arousal and consequently the cognitive load on participants [175]. More specifically, they point out that the HRV is a good measure for cognitive load as its relationship with stress and engagement has been proven [155]. Since cognition and affect are closely related [19, 30], it is important to account for changes in both states to provide a more accurate adaptation and a highly personalised cognitive training.

Another important aspect to achieving positive effects on cognitive skills from video game playing is the overlap of cognitive demands. Transfer of cognitive benefits is more likely when the assessment tasks and the training video game share similar cognitive resources [47]. Oei and Patterson [110] revealed that, in order to achieve any transfer effect playing action video games, it is important that both game and task share a common principle, specific to the skill trained within the game. Therefore, in order to evaluate the benefits induced by game playing, it is fundamental to use the right assessment tools and tasks. Mishra et al [99] reviewed the methodology used in the scientific community for gaming-induced benefits on attention and WM, in order to create more effective game-based training. Although they do not make an implicit distinction between simple and complex span tasks [147], they selected the n-back [70] and the Operation Span Task (OSpan) [162] as the most common and useful tasks to assess the effects of gaming-induced benefits on WM.

Anguera and Gazzaley [8] argue that there are several factors neglected in the design of cognitive video games, such as the role of fun or motivation to play. As reviewed earlier in this chapter (see section 2.3), it is well known that fun and motivation are very important factors to achieve successful results both in learning [91] and cognitive training [160]. For this reason, Anguera and Gazzaley [8] propose a new hybrid video game design that mixes the approach of cognitive tasks and entertainment video games, requiring "*a close collaboration between the video game industry and cognitive neuroscientists to create the next generation of cognitive training tools*". This approach was successfully applied to their game *NeuroRacer* [7], which also uses adaptation algorithms to adjust the difficulty level to each individual player. Thereby, CLVGs can sustain optimal levels of fun, motivation and engagement by dynamically adapting the difficulty level, which can have positive effects on the player's cognitive performance [103].

Finally, the proliferation of new technologies like VR or motion tracking devices brings the possibility to create more immersive and engaging experiences. Rolle et al [134] have tested the suitability of these technologies for cognitive science research, specifically of iPad and Kinect. According to their conclusions, the interactivity of both devices compared to the traditional PC, raises the level of attention and engagement with the task given by increasing the motivation to maximize the performance. These results encourage the direction of this research to explore the potential benefits of VR in the investigation of cognitive training, which can be used to create virtual environments in a more realistic and immersive way as well as allowing a higher control of the stimulus presented [98].

2.5 Summary

This chapter reviewed existing work relating to the study of video games for cognitive training, affective gaming and the link between affect and cognition. Video games are known to be fun, motivating and sometimes challenging for the player's skills [88]. Important gaming aspects like immersion or engagement have been frequently ignored in the design and development of video games for cognitive training. Recent studies [8, 7] propose the incorporation of adaptive mechanisms that dynamically adjust the game's difficulty level in order to provide appropriate challenges. These adaptive video games, also called CLVGs, often alter the game depending on the player's performance, ignoring the experienced affective states. Since arousal and valence can have important effects on our cognition [19] and video games often involve emotional processes [24], some researchers suggest that CLVGs should consider not only the players' performance but also their affective states [98]. Thus, CLVGs should provide a tailored cognitive training that challenges the player's cognitive skills and keeps him or her in an optimal affective state to achieve real cognitive improvements and transfer effects.

Although the negative effects of excessive video game playing such as addiction or aggressive behaviour have been extensively studied [60, 73], this is out of the scope of this research. As the incorporation of adaptive mechanisms that try to keep players in a flow state can increase the chances of game addiction [40], game designers should be cautious to avoid elements that could induce addiction. Thereby, this research is focused on the positive effects of games designed for wellbeing, especially for cognitive training.

This research investigates the effects of VR gaming on the player's WM performance and affective states, as well as the benefits of arousal and valence on WM. An adaptation engine is created to dynamically adjust the difficulty of a VR game depending on the player's affective states and performance. The adaptation aims to keep the player in an optimal affective state to improve the WM performance. The following chapter documents an initial pilot study where some of the methods reviewed in this chapter are tested. In particular, physiological and behavioural signals of two co-located players playing in three play modes (solo, competitively and collaboratively) are investigated to assess its validity to measure the players' affective states.

Chapter 3

Pilot Study: Physiological and Behavioural Differences in Competitive, Collaborative And Solo Gaming

This chapter documents a pilot study that explores the affective, physiological and behavioural differences between sets of two co-located users playing a Wii video game in three play modes: solo (one-vs-computer), competitive and collaborative. This pilot study investigates how collaborative and competitive play modes affect the players' interactions and whether these play modes can be assessed by measuring players' physiological and behavioural signals. Player's physiological signals (cardiac activity and skin conductance) were measured and non-verbal behaviours observed to infer their affective states. Players also self-reported their emotions as well as levels of immersion, engagement and enjoyment after each play. The results reported lead this research towards the use of affective states for real-time video game adaptation depending on the player's affective states, doing a real-time analysis of their physiological and behavioural signals.

A Wii console was chosen due to its novel gesture-based interaction controller. Previous studies that have looked at affective states in competitive or collaborative play modes have used standard controllers like gamepads, joysticks or keyboards. According to Nijhar's findings, gesture-based interaction controllers can increase the player experience and engagement [108] This type of interaction also allowed the capture of user's non-verbal behaviour since they had to play standing up, which made their body positions, spatial behaviour and gestures more obvious and easy to analyse visually. The game played was *Boom Blox: Bash Party*, a physics-based puzzle video game developed by Electronics Arts[1] and designed by the film director Steven Spielberg. This game was chosen as it offers the three different play modes of interest: solo, competitive and collaborative.

Previous studies have demonstrated that it is possible to infer the player's affective state just looking at physiological or behavioural indicators [158, 139, 142]. Mandryk and Inkpen [67] examined the physiological signals of two co-located players in solo and collaborative play modes. Their results showed that the measured physiological signals were good indicators of the reported player experience. In a similar study, Chanel, Kivikangas and Ravaja [36] assessed players' physiological signals playing competitive and cooperatively. They concluded that physiological cues were more correlated in competitive than cooperative play, although this could be explained by the intensity of the social interactions rather than the competition between players. Thus, the pilot study presented in this chapter extends the work by Mandryk and Inkpen [67] analysing physiological and behavioural differences between solo, collaborative and competitive play modes.

In this chapter, section 3.1 describes the design, tasks and procedure of this pilot study. The physiological and motion sensors employed, as well as the questionnaires used, are explained in section 3.2. This section also documents the pre-processing and feature computation of the data collected. Following this, section 3.3 explore the data collected, analysing it at five different levels: i) between play modes, ii) across play modes and pairs, iii) within pairs, iv) between play modes at individual level, and v) behavioural. Finally, section 3.4 discuss the results of the previous section and outline the major findings and limitations of this pilot study.

3.1 Research Design

This section describes the methodology used in this pilot study where eight participants in pairs played a Wii video game in three play modes: collaborative, competitive and solo. A mixed factorial design was used, where between and within-subjects variables were analysed.

[1] Electronic Arts, also known as EA, is a video game developer, marketer and publisher company based in USA: http://www.ea.com/

3.1.1 Tasks and setup

Participants played a video game for Nintendo's Wii[2] console called *Boom Blox: Bash Party* (Fig. 3.1). The game consists of knocking down a structure made of blocks by throwing balls against it using the Wii remote controller as if it were the ball itself. Each block has a number drawn on it indicating how many points it gives when knocked down. Participants could rotate the camera to hit the blocks from different positions. The goal of the game is to get as many points as possible. The players had to play *Boom Blox: Bash Party* in three play modes. The duration of each play lasted between 4 and 7 minutes and the play order was randomised for every pair to avoid bias. The three play modes were:

- Solo: players play alone and get points by knocking down the structure. While one member of the pair was playing this play mode, the other player had to wait outside the room to avoid any bias or communication between them.
- Competitive: two players compete against each other to get as many points as possible knocking down the blocks. An important fact to consider in this play mode was the rotation of the camera. Due to the simultaneous interaction of the two players, both could rotate the camera at the same time, adding a level of complexity since it could be used to distract the opponent.
- Collaborative: the goal of this play mode was not to collect as many points as possible but to play together to break the structure with as few throws as possible. The number of throws were counted for both players, so they had to communicate and think about how to do it the most efficient possible way. Before the game started, participants were informed of how many shots were needed to achieve the gold, silver or bronze medal.

3.1.2 Procedure

The experiment was held in a quiet room with a 40 inch television. First, participants were asked to read and sign the ethics approval as well as the study's instructions. The experimenter explained how to put the electrocardiogram (ECG) electrodes (see Fig. 3.2) in the chest and placed the sensor device in player's waist with an elastic band. The experimenter set the Galvanic Skin Response (GSR) electrodes in the hand holding the game controller with another

[2]Nintendo's Wii console: http://www.wii.com/

Figure 3.1: Screenshot of Nintendo's Wii game *"Boom Blox: Bash Party"*

elastic band on the player's wrist to sustain the sensor. Then, players were asked to sit down and relax for a couple of minutes in order to record their resting physiological signals, which would be used as a baseline to normalise the physiological responses during the study.

Once the physiological baselines were recorded, participants were asked to fill the pre-experiment questionnaire. Then, participants practiced with the game for 5 minutes and were told the order of the play modes assigned. After playing each play mode, participants were asked to sit down and fill a post-condition questionnaire to report their levels of engagement, immersion and enjoyment. The time spent completing this questionnaire allowed participants' physiological signals to go back to their baseline level. Once the three play modes were finished, sensors were removed from players and they were asked to complete a post-experiment questionnaire about their overall experience in the study.

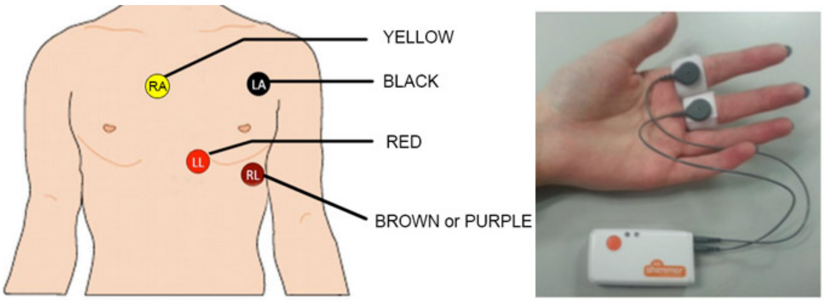

Figure 3.2: Position of electrodes for ECG (left) and GSR (right) Shimmer sensors

3.2 Data Collection and Pre-processing

This section describes the methodology employed to collect self-reported, physiological and motion data, as well as the pre-processing and feature extraction of this data.

3.2.1 Participants

Eight players (four pairs) took part in the study with a mean age of 30.88 (SD: 4.28). A similar sample size of 10 pairs was used by Mandryk and Inkpen in a previous study [67]. Three of them were female and five male. Although none of participants played Wii frequently, four of them played video games between 3 and 5 hours in the last week and only one more than 5 hours. The most common platform to play video games was smartphones. Half of the players reported to prefer playing video games alone, three collaboratively and only one competitively. Participants were recruited via email or word of mouth, trying to involve people with different backgrounds, ages and sex. No prior experience was required to participate in this study, except both players within a pair had to know each other before the experiment to increase the chances of collaboration between them [32].

3.2.2 Sensors and Questionnaires

Self-reported and physiological data was recorded from all participants in every play mode. Before starting the experiment, participants were instructed about the data measures, sensors used and how they should wear them. Table 3.1 summarises the data gathered during the study as well as the sensors employed and features extracted.

Two physiological sensors were used with each participant. A Shimmer[3] ECG sampled at a rate of 512Hz, measured the heart's electrical activity. Participants were instructed about how to place the four ECG electrodes in their chest (see Fig. 3.2). The second physiological sensor was a Shimmer GSR, which measures the electrodermal activity of the user's sweat glands. This activity is measured by passing a low voltage across two electrodes attached to the user's index and middle finger (see Fig. 3.2). The electrodermal activity varies with the state of the sweat glands of the skin, which are normally associated with stress and anxiety as they are related to the sympathetic nervous system, being a good indicator of emotional arousal [41]. This sensor was placed in the hand holding the game controller and sampled at 51Hz. The GSR also had an accelerometer incorporated to record the movements of the hand holding the

[3]http://www.shimmersensing.com/

controller. Due to individual differences in physiological signals, baseline activity levels were recorded at the beginning of the study for all sensors to normalise these differences. Both sensors broadcasted the data wirelessly to a Windows laptop, where the Multi Shimmer Sync software recorded them.

Table 3.1: Objective (continuous) and subjective (self-reported) data recorded.

Measure	**Sensor / method**	**Features**	**Type of data**
ECG	Shimmer ECG sensor	Heart Rate (HR): mean & SD. Inter-Beat Interval (IBI): mean. Heart Rate Variability (HRV): Root Mean Square of Successive Differences (rMSSD)	Quantitative
GSR	Shimmer GSR sensor	Skin Conductance Level (SCL): mean & SD. Skin Conductance Response(SCR): mean & SD.	Quantitative
Motion	Accelerometer	Number of throws (peaks), Quantity of motion, highest peak (velocity throw).	Quantitative
Video	Front-facing camera	Gestures, posture (body position), spatial behaviour, number of gazes, positive and negative facial expressions...	Quantitative and qualitative
Self-report	PRE, PPQ & POST	Engagement, immersion, frustration, stress, enjoyment, effort, boredom.	Quantitative

Three questionnaires were designed using 5-point Likert scales (see Appendix A). A pre-experiment questionnaire (PRE) was given at the beginning of the study, asking demographic (i.e.: gender, age, occupation...) and gaming habit questions such as number of hours spent playing video games in the last month or preferences of video games types. A second questionnaire was given to participants after they finished each play mode. This post-play questionnaire (PPQ) contained questions extracted from two validated questionnaires: the Game Engagement Questionnaire by Brockmyer et al. [29], and the Immersion Questionnaire by Jennet et al. [69]. In this questionnaire participants reported their levels of enjoyment, effort, engagement and immersion overall and with his or her partner. Questions were randomised to avoid any bias due to repetition of the questionnaire after each play mode. Finally, a post-experiment questionnaire (POST) was completed at the end of the study, where participants reported their preferences of each play mode in terms of the most fun, boring, frustrating, etc. Participants also reported their overall engagement and enjoyment levels for each play mode.

Finally, a video camera placed next to the monitor displaying the game recorded the study for later observational analysis and qualitative data extraction. Since participants had to play standing up, these recordings were important to look at their non-verbal behavioural cues such as postures, gestures or facial expressions.

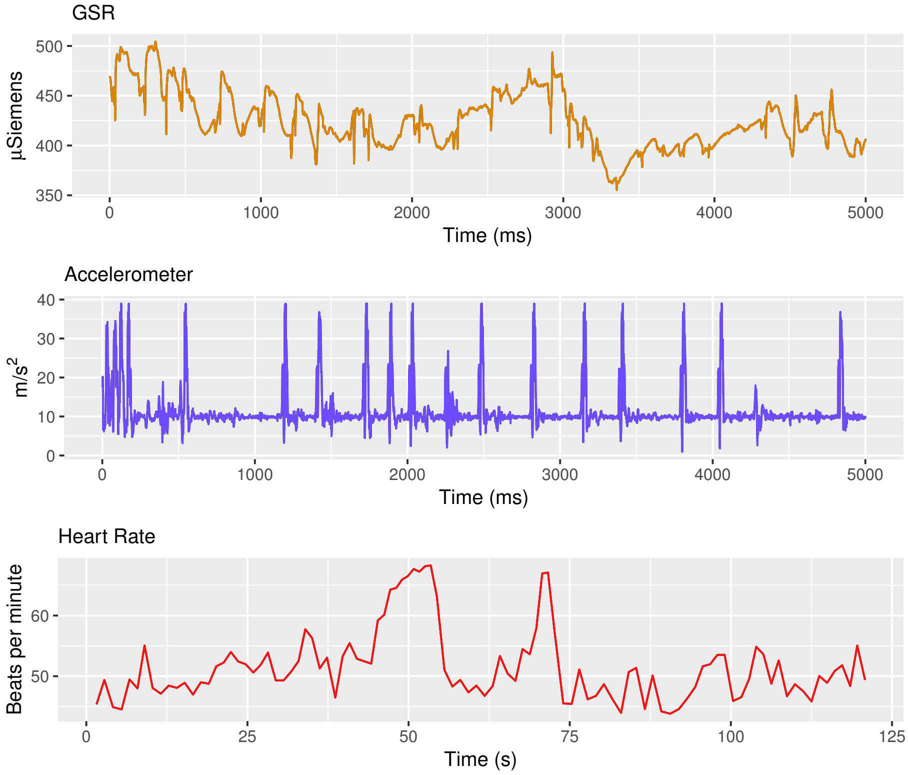

Figure 3.3: Sample data of GSR (top), accelerometer's magnitude (middle) and HR (bottom).

3.2.3 Data pre-processing

All the data gathered from the ECG, GSR and accelerometer during the study was imported into MATLAB. Although a sampling rate of 512Hz and 51Hz were set up for the ECG and GSR sensors respectively, all the sensors' data was recorded at 51Hz due to problems with the software recording all sensors' data (Multi Shimmer Sync).

Data from all sensors was plotted to check if it was correct. The GSR data was extremely noisy since it was placed in the hand holding the controller, which was constantly moving and shaking. As shown in Figure 3.4, the peaks in the GSR signal match the accelerometer's peaks, which correspond with the throwing gestures made while playing. Different filters such as lowpass or moving average filter were applied to remove the noise introduced by the hand movements, making the data considerably smoother although it was still too noisy to use in this analysis so it was discarded. Previous works in affective gaming research have unadvised using GSR for fast paced games that require rapid movements or fingered dexterity [154]. Thus, GSR is suitable only for games that induce relaxation as the hand with the GSR sensor attached needs to be still at all times.

ECGTools[4] was used to analyse the ECG data and extract the R-peaks, which corresponds to individual heart beats. The distance between consecutive R-peaks (also called R-R intervals) was calculated to find the Heart Rate (HR) values per second. The R-R intervals were also used to compute the Inter-Beat Intervals (IBI), which represents the distance in milliseconds between individual heart beats. The HR and IBI are closely related since a higher HR implies a smaller the time between R-R intervals, which mean smaller IBI values.

3.2.4 Feature Computation

Using the IBI calculated from the R-R intervals, different Heart Rate Variability (HRV) features can be extracted. HRV measures the variation of the frequency of heart beats over time. Since there are different ways to measure the HRV, both in the time domain (AVNN, SDNN, rMSSD, etc) and frequency domain (LH, HF), the Root Mean Square of Successive Differences (rMSSD) was calculated as it is one of the most common measures in the time domain [150, 153]. While HR has been demonstrated to be associated with emotional regulation and arousal [158], HRV is linked with stress and mental efforts like engagement [10][67]. The mean values of HR, IBI and rMSSD were calculated for each participant and play mode. Moreover, the continuous IBI data was interpolated to

[4]http://www.ecgtools.org/

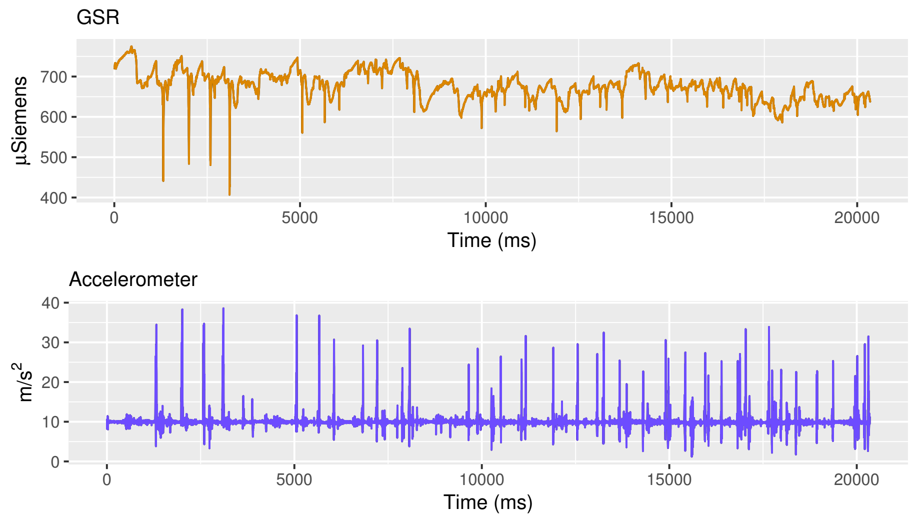

Figure 3.4: GSR (top) and accelerometer magnitude (bottom) during the competitive condition.

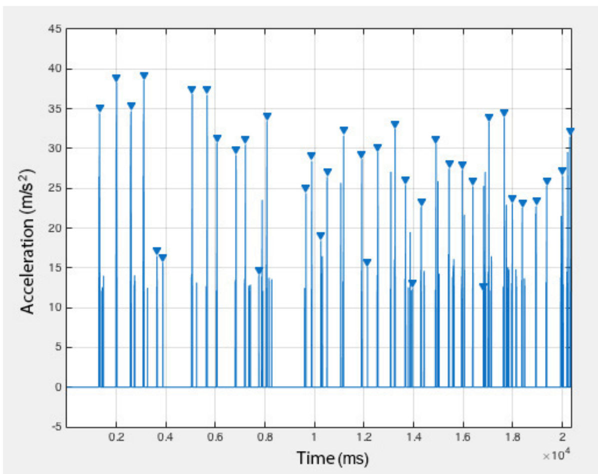

Figure 3.5: Accelerometer peak detection. Each triangle corresponds with one throw.

have an evenly spaced time series in order to be able to do a continuous analysis over time.

The acceleration magnitude (change over time in velocity, m/s^2) of the hand holding game controller was computed using the absolute value of the accelerometer's X, Y and Z axes. This acceleration magnitude was not gravity-free. The mean and standard deviation of the acceleration magnitude, hereafter referred to as Quantity of Motion (QoM), were calculated. Using a peak detection (see Fig. 3.5) algorithm in MATLAB, participants' throwing gestures were identified. Thereby, the number of throw gestures of each participant per play mode were calculated. In order to assure all throw gestures were detected, the parameters of the peak detection algorithm were adjusted until the peaks detected corresponded with all the observed throws in the video recordings.

A qualitative analysis of the videos was carried out. Since some recordings were a bit blurry, it was difficult to use Computer Vision techniques to accurately detect participant's facial expressions to perform an automatic emotion recognition analysis. Therefore, video recordings were manually annotated to determine the predominant facial expressions in each play mode as well as the non-verbal behaviours such as gestures, postures or spatial behaviour (i.e.: moving around the room). This analysis investigated how participants express the experienced affective states through their body movements and facial expressions, distinguishing between positive, negative and neutral states. Finally, the data gathered from the PRE, PPQ and POST questionnaires was imported into the statistical analysis software SPSS (Version 30) since it did not need any pre-processing.

3.3 Analysis and Results

The aim of the analyses reported were to investigate the affective, physiological and behavioural differences between play modes. Three levels of analyses were undertaken: i) overall, with all participants; ii) within pairs; and iii) individual. This allowed the analysis not only to investigate general trends over all participants, but also a more detailed analysis looking at individuals and the relationships between pairs. Between and within-subjects analyses were carried out to investigate the effects of the aforementioned play modes on the players' experience and affective states.

3.3.1 Analysis between play modes

Four out of the eight players reported in the POST questionnaire to have enjoyed the collaborative mode the most, whilst three preferred playing competitively and only one the solo mode. Five participants declared the competitive mode was the most stressful, while two felt more stressed playing collaboratively and only one in the solo mode. Players reported to be equally engaged with their partners during the collaborative (M:3.88, SD: 0.99) and the competitive play mode (M: 3.25, SD: 1.16). No significant differences were found in the immersion level.

In order to analyse whether there was significant differences between play modes in the continuous features, various paired t-tests were carried out, specially between the competitive and collaborative play modes (see Table 3.2). One of the players who reported neither enjoyment nor engagement in any of the play modes and whose physiological signals did not vary much across play modes, was removed in the HR paired t-test. The mean HR and the mean IBI showed significant ($p < 0.01$) statistical differences between the competitive and collaborative modes, with a mean variation of 10.67 in the HR (SD: 4.09) and -100.24ms in the IBI (SD: 63.35). This indicates that participants experienced a higher HR of 10 Beats Per Minute (BPM) average in the competitive play mode. Note that the high standard deviation of HR and IBI is due to individual physiological differences between participants. For example, the mean HR of two players in the competitive mode was 85 and 65, while in the collaborative mode it was 81 and 55 respectively. This means that although there are significant differences in the mean HR between competitive and collaborative modes, there is an important deviation from the mean depending on individual physiological differences .

The accelerometer showed some significant ($p < 0.01$) differences, particularly in the number of throws and the mean QoM. The statistical difference of number of throws had a mean of 20.85 (SD: 7.01), which means players made an average of 20 throws more in the competitive play mode than in the collaborative. This high difference was caused not only by the nature of the collaborative condition where players had to make as few throws as possible, but also due to the turn taking strategy all pairs took while playing together, even though participants were informed they could play at the same time. Moreover, the mean QoM was higher in the competitive mode by 1.16 (SD: 0.54). This result is closely related to the number of throws, since more throws means more motion of the controller. Finally, no significant differences were found in the fastest throw (i.e highest peak in accelerometer) between play modes.

3.3.2 Across play modes and pairs

This analysis focused on the relation between continuous and self-reported data of all players. The aim of this analysis is to explore what continuous or self-

Table 3.2: Paired T-tests between play modes

Param.	**Mean**	**SD**	**t(df)**	**Sig.**
Mean HR	10.67	5.25	5.38 (6)	.002
Num. throws	20.75	7.50	9.03 (7)	.000
Mean QoM	1.07	.56	5.35 (7)	.001
Fastest throw	.50	.78	1.83 (7)	.109

Note: SD = Standard Deviation.
df = degrees of freedom.

Table 3.3: Spearman's correlations between continuous and self-reported data

Param.	**Depend. Var.**	**Rho**	**Sig.**
Mean QoM	Norm. Mean HR	.413	.052
Mean QoM	Norm. Mean IBI	-.462	.053
Effort	Norm. Mean HR	.584	.001
Engage w/ Partner	Norm. Mean HRV	-.535	.001
Enjoy w/ Partner	Norm. Mean HRV	.265	.211
Enjoy w/ Partner	Norm. Mean HR	.122	.571
Flow	Norm. Mean HR	-.157	.497

reported variables would be useful to infer the player's affective states and behaviours. Prior to this analysis, the physiological data of each player was normalised to mitigate individual differences. Each player's ECG data was normalised according to his/her own baseline, recorded at the beginning of the study. In order to normalise the physiological data of each play mode, the mean of the baseline was divided by the mean of each mode, getting the percentage of increase for a particular play mode. For example, if the mean resting HR of one player was 73 BPM and the mean HR for this same person was 96 BPM in the competitive play mode, it can be said that the HR increased by 31% in that particular play mode.

$$Normalised = \frac{PlayMode}{Baseline}$$

Once all the physiological data was normalised,rho= it was correlated with objective (QoM) and subjective variables such as effort or enjoyment (see Table 3.3). The mean QoM had a significant moderate positive correlation with the normalised HR (rho=.413, $p < .05$) and, at the same level, was negatively correlated with the IBI (rho=-.462 $p < .05$). This correlation between QoM and HR is probably related to the significant correlation (rho=.584, $p < .01$) of the mean HR with the self-reported effort in the post-play questionnaire. These correlations are meaningful, since the BPMs increased with the required movement and effort needed to achieve a good performance.

The mean of the normalised HRV rMSSD showed a strong significant but negative correlation with engagement with the partner (rho=-.535, $p < .01$). This significant negative correlation demonstrates that HRV is lower when the player is more engaged. This finding is in line with previous studies that demonstrated HRV decreases with mental effort [67]. When the player is more engaged with the game, the body tends to be more relaxed and the heart activity settles down without much fluctuations. Thereby, as a result of higher engagement levels, participant's attention increase, which decreases the player's HRV.

Aside from these significant correlations, other interesting but non-significant correlations were also explored. These results were not statistically significant probably due to the small number of participants in the experiment. The normalised mean HR and HRV showed a low correlation with the reported enjoyment with the partner, being higher for the HRV rMSSD (see Table 3.3). Moreover, the level of flow had a small negative correlation (rho=-.157, p=.497) with the normalised mean HR. This can be related to the negative correlation of engagement with partner and HRV, as both engagement and flow refers to the focus or concentration levels. Thus, a higher engagement or flow state would make the heart activity more steady and, in this case, also slower, although this would depend on other factors like the amount of exercise made while playing [67].

3.3.3 Analysis within pairs

This section compares the behavioural and physiological responses between the players within each pair, investigating the correlations in the continuous signals of the two players. Due to the intrinsic auto-correlation of the ECG signal, it is not possible to perform a simple cross-correlation with this data as it is biased [15]. One way to overcome this problem is to make a 1 second interpolation of the HR values in order to have an evenly spaced continuous data. Then a non-overlapping window of 3 seconds was generated for every participant and play mode. Once the data was windowed, it was possible to perform a normal Spearman's correlation between members of one pair in each play mode separately.

Table 3.4: HR correlations between pair members.

	Pair 1	**Pair 2**	**Pair 3**	**Pair 4**
Comp.	.509**	.165**	-0.066*	.200**
Collab.	.021	.034	.061	.022

Note: Significance of Spearman's rho: **p < .05; **p < .01.*

As shown in Table 3.4, all players experience a significantly higher correlation in HR when playing competitively than collaboratively. The higher correlation of pair 1 during competitive mode might indicate an emotional contagion between players, as player 2 of this pair was the only participant who reported to prefer playing competitively. Chanel, Kivikangas and Ravaja [36] suggested that emotional contagion is one of the factors that could explain physiological compliance in competitive gaming. However, the low (and in some cases negative) correlation in the collaborative play mode can be again explained by the turn-taking strategy followed by all pairs. In the collaborative mode, whilst one player experienced high levels of arousal, the other player was more relaxed waiting for his or her turn. For example, pair 3 in competitive play mode have a very small negative HR correlation due to the lack of engagement, immersion and even enjoyment of one of the players as reported in the POST questionnaire. Therefore, player 1 was more aroused than player 2 as their HR differ considerably.

3.3.4 Individual analysis between play modes

Since there was an interest in how players experience each play mode, participants' physiological and self-reported data in each play mode was examined. As shown in Table 3.5, the strong correlation of effort with engagement and immersion in the competitive mode indicates that higher effort leads to higher levels of immersion and engagement. However, this is not true for the collaborative mode, although these results can be affected by the turn-taking strategy, which required less effort.

The normalised mean HR in the solo and collaborative modes were also correlated at an individual level, looking for relations in the physiological responses in these play modes. This analysis evidenced a very significant and strong correlation (rho=.929, $p < .01$). This indicates that when a player is relaxed playing the solo mode, s(he) will experience a similar level of arousal when playing collaboratively.

The mean HRV rMSSD was correlated with the self-reported "*fun with player's partner*" (Table 3.5). Although the result was not significant for the collaborative mode, it almost was for the competitive mode. These results suggest that the fun level when playing competitively is associated with a higher variability in the heart activity and therefore a higher degree of stress and frustration, since a high HRV is associated with these affective states [67].

Table 3.5: Spearman's correlations between play modes at individual level

Play Mode	**Param.**	**Depend. Var.**	**Rho**	**Sig.**
Competitive	Effort	Engagement	.756	.030
	Effort	Immersion	.571	.139
	HRV	Fun w/ partner	.639	.088
Collaborative	Effort	Engagement	-.103	.808
	Effort	Immersion	.130	.759
	HRV	Fun w/ partner	.041	.923
	Mean Norm. HR Solo	Mean Norm. HR Collab.	.929	.001

Figure 3.6: Participants playing competitive (left) and collaboratively (right).

3.3.5 Behavioural analysis

The analysis in this section focuses on the video observations of the facial expressions, gestures, postures (body positions) and spatial behaviour of players. Spatial behaviour or movement can be described as the activity of one individual moving through the surrounding environment (the room). Participants' facial expressions were labeled into 3 groups: positive (happy), negative (frustrated or angry) or neutral. While positive facial expressions were annotated when participants smiled or laughed, negative facial expressions were annotated when participants were frowning or pressing the lips together. Neutral expressions were collected when no facial expressions were displayed.

Each recorded video was divided into three equal parts, where the predominant facial expressions for each part was annotated. The annotation was carried out by a trained experimenter who counted the number and duration of the facial expressions in each part. The most common expressions in the competitive mode was negative as the players tried to win but not always got the expected results (getting stressed and even angry). This can also be explained by the higher levels of effort reported in the competitive mode. Positive facial expressions were also present in the competitive play mode, usually appearing at the end of the game when both players relaxed and talked about their performance. Some participants had recurrent 'specific' facial expressions such as biting their lower lip, sticking out the tongue or frowning, which displayed their frustration or engagement. On the other hand, the collaborative play mode elicited a higher number of positive facial expressions and laughter, whilst neutral faces were the most frequent in the solo mode.

When labelling postures, gestures and spatial behaviour of players, their reactions and behaviours over a whole play mode were observed. Overall, players had a more relaxed behaviour and body posture during the collaborative and solo play modes than in competitive mode (Fig. 3.6), displaying a greater spatial movement and more gestures such as head nods or moving arms around their body. Players changed their postures (body position) more often in the collaborative play mode (normally after each throw) and had more social interactions, mutual glances and conversations, not only due to the turn-taking strategy discussion but also talking about the game or their performance. Conversely, in the competitive play mode, players were more static, barely moving their body or legs, and rarely speaking to each other until the game was over.

3.4 Discussion

The significant correlation of HR between players during the competitive mode, plus the significant mean HR difference compared to the other two modes, demonstrate a clear arousal increase when playing competitively. The strong correlation of the normalised mean HR in the solo and collaborative modes show that the arousal level in these modes are related. In other words, players are likely to experience the similar levels of arousal when playing alone than when playing collaboratively. HRV is also an interesting cardiac feature to measure engagement [158], evidenced by a significant negative correlation with the self-reported engagement with partner.

However, HR must be interpreted carefully in gaming. While an increase in HR, caused by the cardiac sympathetic activity, is associated with affective arousal, a slow HR inflicted by the cardiac parasympathetic activity is related to attentional engagement [126][161]. Since video games can evoke both states simultaneously, HR must be interpreted carefully when using it to measure arousal in games [127]. Although skin conductance can be a good and unambiguous indicator of arousal [139], it is inappropriate for studies like this one where participants are constantly moving their hands [154].

Aside from the aforementioned physiological manifestations of arousal, video observations also revealed that competitive play mode evoked a tense behaviour in players. This was manifested through static postures (Fig. 3.6), certain facial expressions such as sticking the tongue out or frowning, and the lack of verbal interaction or spatial movement. On the other hand, players were more relaxed when playing collaboratively, showing more positive facial expressions. The number of social interactions among players was higher during the collaborative play mode, not only with non-verbal behaviours such as postures, gestures or glances, but also talking more about the game, strategy and performance.

Participants' non-verbal behaviour should be interpreted carefully. A greater number of negative facial expressions or negative affective states in a particular play mode does not implicitly mean a decrease in players' enjoyment. As remarked by Lazaro et al [84], different players can experience affective states in different ways, normally depending on their motivation to play. In a similar manner, Fairclough [55] argued that certain levels of frustration might be tolerated for short periods of time as the player is engaged and challenged. For instance, players looking for challenges might enjoy negative affective states such as tension or frustration. Thus, if a player's motivation is to achieve the overall best score, s(he) may enjoy the competitive play mode and the associated tension it could elicit [108].

Finally, it is interesting to mention the turn-taking strategy all pairs followed on the collaborative mode. Even though players were informed that they could play at the same time, independently from the other player, all pairs followed a turn-taking strategy. Interestingly, these turns did not strictly alternate from player to player as sometimes one player threw more than 2 times in a row. This is probably indicative of some sort of emergent friendship dominance, although this is out of the scope of this study.

3.4.1 Findings

The results reported in this chapter can be summarised in three findings at three distinct levels of analysis:

1. The analysis at the overall level demonstrates that participants experienced a higher levels of arousal and tension when playing competitively than in collaborative or solo play modes. Significant differences in players' HR in the competitive mode support this finding. Moreover, the observed facial expressions, postures (body positions) and gestures evidence higher tension and frustration levels in the competitive mode. Similar results have been found in previous studies examining players' physiological signals in video game playing [67]. The overall level analysis also showed evidences that HRV can be a good indicator of engagement when playing video games. The significant negative correlation between HRV rMSSD and the self-reported engagement suggest that high levels of engagement may reduce the HRV. Low levels of HRV are related to sustained attention [153], memory performance and mental workload [67]. However, the

HRV analysis in this study was limited to one feature (rMSSD), ignoring other HRV features both in the time and frequency domain. More research needs to be done in gaming to further confirm the relationship between HRV and engagement.

2. The within pairs level analysed the relationship of continuous measures between players within the same pair. The analysis revealed that players' HR are significantly correlated during the competitive play mode but not during the collaborative mode. This may be explained by an emotional contagion between players, one of the factors suggested by Chanel, Kivikangas and Ravaja [36] to explain physiological compliance in competitive game playing.

3. The individual level analysis revealed that players experienced similar levels of arousal during both solo and collaborative play modes. The strong significant correlation of participant's HR in these two play modes demonstrate that players show alike activation levels when playing alone and collaboratively with a friend. This result contributes to the findings of previous works investigating physiological signals in single and multiplayer games [36, 67]. Furthermore, the significant correlation between the self-reported effort and engagement in the competitive mode suggest that participants were more engaged when they put more effort in winning while competing, but not while collaborating.

These findings can be useful for researchers investigating the effects of video games in the affective states of two co-located players. Thereby, this research support previous works claiming that physiological responses can be a good indicator of the player's affective states [67, 158, 106] However, it is important to note that different results may be obtained using different games or interaction controllers. An interesting research direction is to explore how this data could be used to enhance player experience in competitive and collaborative gaming. Although this study has mainly focused on arousal but not valence, it is important to consider both dimensions to infer the player's affective state. An adaptation engine could use the inferred affective states to adjust certain game elements or parameters depending on the player's emotions in order to keep the player in an optimal affective state.

3.4.2 Limitations

An important limitation of this study was the small number of participants. Only four pairs (eight participants) took part, which might explain why some results were not significant and the low statistical power of these analyses. However, the results described are indicative of the affective states and behaviours players manifest when playing in different play modes such as collaborative, competitive or solo. For this reason, further research needs to be done to confirm the statistical validity of the reported results.

Another limitation were the problems with the GSR data due to placement of the sensor in the hand holding the Wii controller. This sensor aimed to measure both the player's electrodermal activity and hand's motion. However, the data recorded was very noisy due to the constant hand motion. Although GSR is a good and unambiguous sensor for measuring arousal [139], it is very prone to noise if it is used in motion [154]. This means that GSR is not suitable to measure arousal in studies like the one here described, where participants have to constantly move their hands.

Finally, due to the intrinsic autocorrelations (correlation with itself) of physiological time series [48], the cross-correlations performed between pair members' HR might not be valid and should be interpreted carefully. Dean and Dunsmuir [48] proposed a method to avoid issues with autocorrelation in human-derived time series such as movement or physiological series. This method consists on two steps: first, the series should be made stationary, making the series with a constant mean and standard deviation. Then, the autocorrelation can be removed following a process known as '*pre-whitening*', which consists on fitting and autoregressive model to one of the series and filtering it with the parameters of the model created. Even though it seems reasonable that participants' arousal levels are correlated with their partner when playing competitively but not collaboratively (due to the turn-taking strategy), the cross-correlations results reported in Table 3.4 should be interpreted carefully. Furthermore, the strong positive and almost perfect correlation (rho=.929, $p < .01$) of participants' HR in the solo and collaborative modes should be further explored as it might be spurious.

3.5 Summary

The pilot study reported in this chapter has analysed the physiological signals and non-verbal behaviours of two co-located subjects playing in solo, competitive and collaborative play modes. Participants played the Wii game *Boom Blox: Bash Party*, which consists on knocking down a structure made of blocks by throwing balls against it using a gesture-based controller. The aim of the study was to assess the usefulness of physiological sensors and behavioural observations to distinguish between the different play modes proposed. ECG and GSR sensors were used to measure participants' physiological signals, as well as an accelerometer to gauge the QoM of the hand holding the game controller. Participants also self-reported their levels of immersion, engagement, effort and enjoyment after each play mode. The results indicated that participants experienced higher levels of HR and arousal in the competitive play mode, whilst similar levels were experienced in the solo and collaborative play modes. In line with these results, video observations of postures, gestures and facial expressions revealed higher levels of tension and more negative facial expressions during the competitive mode. No significant differences were found in the immersion levels, although the self-reported effort and engagement correlated significantly only in the competitive mode, which means that putting more effort to win the partner entails higher levels of engagement.

Following the findings and insights gained from this study, certain measures were selected according to their meaningfulness and reliability to assess participant's affective states. As suggested by previous research [158] [41] and demonstrated in this pilot study, HR is an effective and well-known measure to estimate arousal levels. The significant correlations of HRV rMSSD with the self-reported engagement also suggested that HRV can be a good estimator of engagement levels. These conclusions motivated the study described in the next chapter, which investigates the effects of valence and arousal on the player's WM performance when playing in two interaction modes: Desktop and VR. It also assesses the WM performance of participants playing a custom-made video game called *Memory Break*. Since video games can have positive effects on the player's cognitive skills [7, 43], there was an interest on exploring the use of video games in VR for cognitive training. Finally, the next study will investigate how affective states can be measured using non-invasive sensors. The Shimmer ECG and GSR sensors used in this study were remarkably invasive for some participants as they had to attach electrodes directly to their chest or hands. Recent advancements in wearable technology allow easier physiological readings in non-invasive and reliable ways [5].

Chapter 4

Study 2: Effects of VR Gaming on WM Performance and Affective States.

The previous chapter reported a pilot study in which physiological and behavioural signals were analysed to assess the affective states and player experience of two co-located players playing a Wii video game in three different play modes: collaborative, competitive and solo. The findings of the study indicated that physiological and behavioural signals can be used to assess the player's affective states. For example, HR and HRV are suitable indicators of arousal and engagement respectively. These results confirm the initial hypothesis that affective states can be inferred using physiological and motion sensors, answering the first research question outlined in Section 1.1.

However, the study of affective states in multi-player gaming contexts implies more challenges than in single-player video games. The interaction between players within the game influences their affective responses, introducing more variability in the results [24]. This was demonstrated in the previous chapter where participants experienced similar levels of arousal when playing competitively, probably due emotional contagion between players [36]. Therefore, it is important to understand first the effects of gaming in a player's affective state before exploring social interactions in multi-player games.

This chapter reports a study that investigates how working memory (WM) performance is affected when playing a VR game, and the effects of valence and arousal in this context. This study also explores the differences in self-reported immersion of participants playing in two interaction modes: Desktop and VR. Thus, this study answers two of the research questions regarding the effects of

VR gaming on WM performance and affective states, as well the effects of the latter on WM. A single-player custom video game called *Memory Break* was created for Desktop and VR settings using a gesture-based interaction. Three difficulty levels (easy, medium and hard) were designed to evoke different levels of arousal while maintaining the same memory load for each difficulty level. Physiological and motion sensors were employed to measure the player's affective states, as well as self-reported data using the Game Experience Questionnaire [66] and the Affective Slider [21].

The structure of this chapter is as follows. Section 4.1 outlines the aims and motivations of the present study. In section 4.1, the tasks and procedure of this research are described, as well as the Automated Operation Span Task, a popular WM capacity test employed throughout the remain of this thesis. The mechanics and interaction of custom video game *Memory Break* are explained in detail in section 4.3. This section also documents a pilot study that tested the game. Moreover, the participants recruited for this study and the questionnaires and sensors used, are described in section 4.4. Section 4.5 reports the statistical methods employed and the statistical analysis performed on the data collected. Finally, section 4.6 discusses the results and implications of this study.

4.1 Aims and motivations

The use of VR in this study is motivated by previous research that referred to VR as an *'affective medium'* due to its ability to evoke and intensify the affective states [132]. VR has brought an increase in the user's immersion and engagement. This new degree of immersion, referred to as *presence*, is reported as the feeling of *'being there'* in the virtual world [132]. Research linking presence and affective states has suggested that higher levels of presence directly influence the vividness and intensity of the emotions which users experience [132]. This study aims to examine the differences in self-reported levels of immersion in Desktop and VR gaming, as well as its effects on player's affective states in both interaction modes.

On the other hand, the influence of affective states in attention and WM have been widely investigated [19, 46, 173]. These two cognitive skills are intrinsically related; attention regulates the incoming information and WM retains it while other cognitive processes are ongoing [147]. Whilst arousal has been demonstrated to enhance attention up to a certain point, after which it has a negative effect, the role of valence is still considered controversial and might be task dependent [19]. Due to the effects of arousal and valence in our cognitive skills, this study investigates the link between affective states and WM performance when playing a VR video game.

Finally, the aim of creating a custom video game instead of using an existing one was to be able to control all game elements. Existing commercial video games can have interactions which are too demanding or game elements that could distract participants [103]. Developing a custom video game allows for strict control of every single game element such difficulty level, timings or graphics.

4.2 Research Design

A within-subjects design was used in this study to compare the effects of game playing in VR and Desktop settings. Participants had to attend two sessions to play the custom video game *Memory Break* in both settings on two different days. The WM capacity of all participants was evaluated using the Automated Operation Span Task test. In contrast to the previous study, participants played the game alone and no qualitative or observational analysis was performed.

4.2.1 Shortened Automated Operation Span Task

The Operation Span Task (OSpan) test is a complex span task that measures WM capacity [56]. Unsworth et al. [162] proposed an automated version of the OSpan for easy administration to save time for both the experimenter and the user. Even this automated version takes around 25 minutes to complete, which is a very long time for the usual constrained time of research studies. Foster et al [56] made a shortened Automated Operation Span Task (AOST) where the experimenter can decide the length of the test by selecting how many blocks to administer. According to their research, 2 blocks of AOST are enough to predict a subject's verbal WM capacity. Mishra et al [99] also suggested using the OSpan to measure the benefits of game playing and game-based training on verbal WM. As these researchers claim, it is important to differentiate between verbal and visual WM.

The test (see Figure 4.1) is divided in 2 phases, the task phase and the recall and feedback phase. The task phase is composed by a sequence of multiple maths problem-answer-letter cycles. First, the user is presented a simple math problem which must be solved as quickly as possible. Once the user has the solution, they then click the mouse and a new screen appears which contains a number and two buttons labeled *True* and *False* show up. If the number presented is the right answer to the math problem shown, the user has to click the *True* button, otherwise they click *False*. After the answer, a letter appears for 800ms, which the user has to remember. Then a new maths problem-answer-letter cycle starts. When the task phase finishes, the user is presented a panel

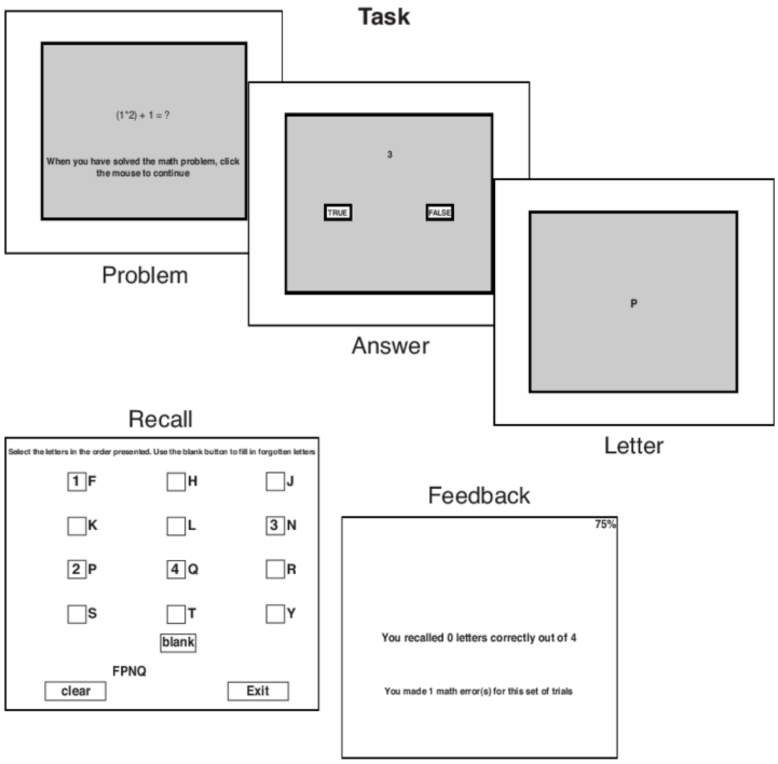

Figure 4.1: Procedure of the Automated Operation Span Task test

with all the possible letters shown where the user has to recall and input the letters previously shown in the exact same order. The user then clicks the *Enter* button on the bottom right corner and feedback is presented informing how many letters were recalled correctly and the number of math errors. The length of each sequence is a random number between 3 and 7 cycles. In one block, each sequence length (3, 4, 5, 6 & 7) can only happen one time. Therefore, each block have a total of 25 letters shown, although the user is not informed about this.

Before the actual test begins, the user is given some time to practice first only with the letters, then only the maths problem-answer and finally with both letters and maths together. During the maths practice, the average time to solve a maths problem for each user is calculated. Users are required to respond the maths problems within 2.5 standard deviations of their average response time [56]. If this time is exceeded, a letter will be displayed automatically without showing the answer screen, and it is counted as a math error. Subjects are asked to maintain at least an 85% of the maths correct in order to use the data reliably. The percentage of maths correct is displayed during feedback on the top right corner of the screen.

4.2.2 Tasks and Setup

The study took place in a quiet room of Queen Mary University of London. This room was chosen due to its soundproof architecture, which isolated participants from the outside noise that could disturb or divert attention from the given task. Participants were anonymised with an ID number. Once the study was completed, participants were compensated with £15 cash for their time and travel expenses.

The study consisted of two sessions, one for each interaction mode: Desktop and VR. Participants were asked to leave one week between sessions and book both sessions at the same time and day of the week to minimise differences that might result from being tested at different times of the day. The structure of sessions 1 and 2 was not exactly the same. Whilst session 1 required participants to take the AOST test before playing the game, participants only had to play the game in session 2.

Participants were randomly assigned to a group to play either on Desktop or VR on their first session. When playing the game in VR, participants played a 3 minutes VR demo called *Blocks*[1] where they could get used to the HMD, the virtual world and the interaction with their virtual hands. This aimed to reduce the novelty effect that a VR experience could cause on affective states

[1] `https://gallery.leapmotion.com/blocks`

[149] as well as to check for any possible motion sickness the HMD could cause.

Before introducing the game, a Polar H7 heart rate sensor was attached to the participant's chest and paired with the HRVLogger [5] app that recorded the participant's heart activity. Once participants relaxed for one minute and the Heart Rate (HR) baseline was recorded, the experimenter asked participants to wear an electromyograph (EMG) sensor (MYO) on the forearm of the hand used to interact with the game. The HR baseline was recorded in both sessions to normalise the HR and HRV of each session independently. Then, the experimenter explained the game mechanics and interaction control.

4.2.3 Procedure

During the first session, participants had to fill in a background and gaming habits questionnaire, followed by the shortened version of the AOST test [56] that measured their WM capacity baseline. In order to make the test time as short as possible, the experimenter verbally explained participants the test's instructions and procedure. This also reduced the fatigue that could cause reading the long test instructions.

After a practice play with *Memory Break*, each participant played each level once in random order. Immediately after playing each difficulty level, participants reported their level of engagement and completed the In-Game module of the Game Experience Questionnaire [66], as well as self-reported their levels of arousal and valence using the Affective Slider [21]. Finally, when all difficulty levels were played, participants completed the Post-Game module of the Game Experience Questionnaire [66] and rated each levels in terms of difficulty, boredom, enjoyment, arousal and focus.

4.3 The game: Memory Break

Memory Break (see Fig. 4.2) is a custom-made game developed in Unity inspired in the mobile game for iOS and Android *Smash Hit*[2]. This game has been used in a similar study by Pallavicini et al [117] that investigated the effects of immersive (VR) and non-immersive (tablet) game playing in player's affective states and immersion. They selected this game (*Smash Hit*) due to its relevance in the world of gaming and the possibility to be played in two interaction modes (VR and tablet). *Memory Break* and *Smash Hit* are infinite runner games, also called endless running[3], where the player is constantly moving forward at a constant speed and it has a very simple interaction such as jumping or shooting. This

[2] http://www.smashhitgame.com/

[3] https://en.wikipedia.org/wiki/Platform_game#Endless_running_game

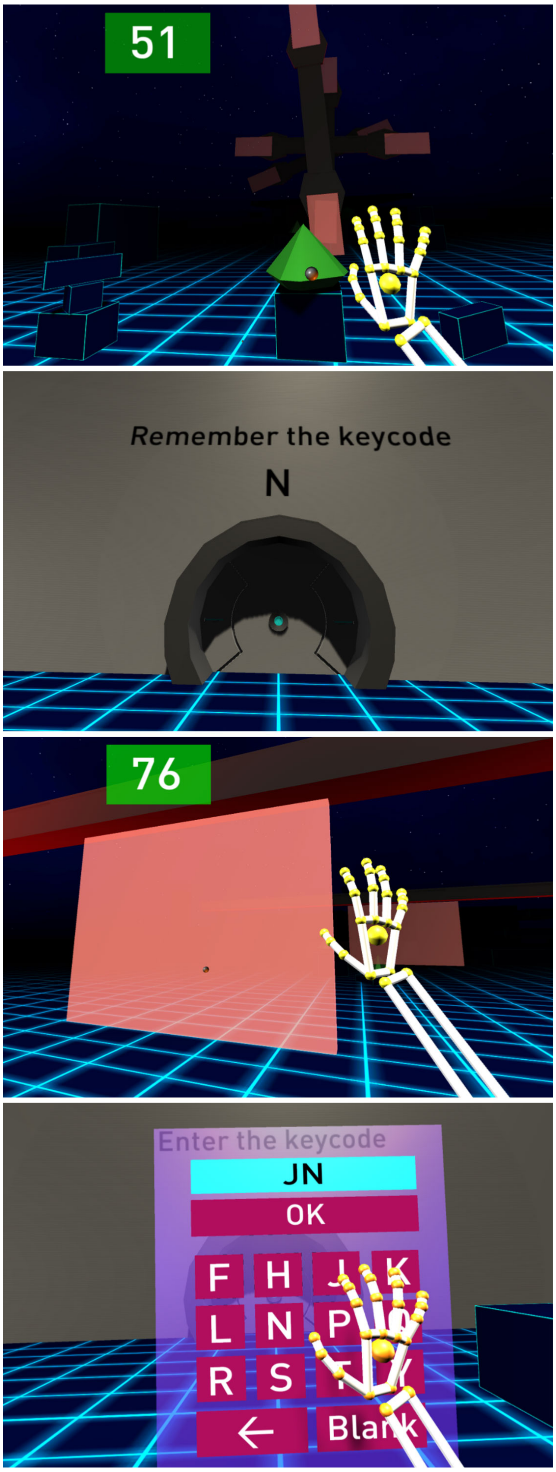

Figure 4.2: The game *Memory Break*

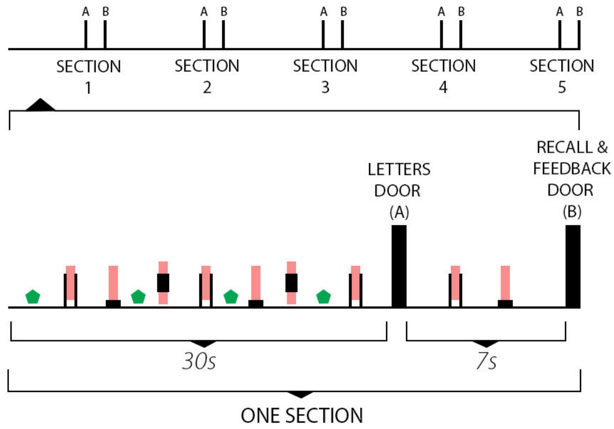

Figure 4.3: Sections of *Memory Break*

type of game, similar to *Project: EVO* [9], was chosen due to its simplicity, low cognitive demands of verbal tasks and easiness to adapt for both Desktop and VR. It requires constant attention and interaction to achieve success since the player does not control the game's pace and has to constantly react to oncoming objects. Thus, *Memory Break* is designed to keep players engaged, immersed and motivated to play.

The goal of the game is to obtain the highest score possible. It consists of throwing balls at different stationary or moving obstacles to successfully pass through without crashing into them. If the player crashes, five points are deducted from the score; one point is also subtracted every time a ball is thrown. In order to get points, the player has to throw and hit the green gems found on the way, which adds 10 or 20 points depending on the type of gem collected. The player has infinite balls to throw and therefore, the score can have negative numbers. This way, the game progress was not dependent on the player's skills, as it would be if the player had a limited number of lives.

Each game play is divided into five sections structured as follows: Every 30s of game play, the game stops at a door where a random sequence of letters appears that has to be remembered. These letters appear one at a time for 800ms, with 500ms gaps between them. The sequence length is randomly selected between 3 and 7 letters, each sequence only appearing once per level. This results in a total of 25 letters per game play, the same number of letters as one block of the AOST test [56] (see 4.2.1). After the sequence is displayed, the doors open and the game continues for another 7s (see Fig. 4.3). These timings are taken from previous research studying WM performance in games and the delayed recall of stimuli [50, 42, 162]. The game then stops again at another door where the player has to recall and input the letters previously shown in the

Figure 4.4: *Memory Break* hand gestures interaction in VR and Desktop settings

exact same order. Once completed, feedback was provided indicating how many letters have been recalled correctly, the doors open and the game continues to the next section. Shooting was disabled and the score was hidden when players stopped in front of a door to avoid distractions.

Three difficulty levels - easy, medium and hard - were designed to evoke three levels of arousal: low, medium and high arousal respectively. The difficulty levels are exactly the same in Desktop and VR. The game's speed is incremented at each difficulty level by 5%. The number of obstacles between doors is also incremented by 2 units, as well as the average number of obstacles per section, being 8, 17 and 34 for the easy, medium and hard level respectively. The duration of each difficulty level is approximately five minutes.

4.3.1 Game interaction

In order to reduce as many dissimilarities as possible between the two interaction modes, the same hand motion tracking device, Leap Motion, was used in both settings. Whilst in VR, this device was attached to the front of the HMD, in the Desktop mode Leap Motion was placed on top of the desktop (see Fig. 4.4). The different physical location of Leap Motion in the two interaction modes led to two different interaction gestures. Due to the different placement of Leap Motion in respect to the player's body, the hand tracking device could not efficiently track the same gesture in both settings. Various gestures were tested, selecting the ones more comfortable and easier to control by unexperienced players. These gestures tried to be comfortable for the players, avoiding any fatigue that could result from repeating the same gestures for a period of five minutes.

4.3.2 Pilot study: Testing *Memory Break*

The game was tested before the actual study to assess whether the three difficulty levels evoked different levels of arousal. Brief informal interviews were carried out to gather suggestions and opinions about the game. Seven participants took part in a pilot study with a mean age of 28.57. Participants reported a mean arousal level of 0.7 (SD:0.16), 0.73 (SD:0.18) and 0.79 (SD:0.16) in the easy, medium and hard levels. As a result, the easy level was modified further in order to make it easier. No significant differences were found on the valence level. These results indicate that three distinct levels of arousal are achieved in the three difficulty levels created, validating the game design. It was important that the hard level was challenging enough but not impossible to play as this could lead to demotivation of the less experienced players.

Some improvements to the game were made according to the participants' suggestions such as adding audio feedback when touching buttons or changing certain colours. None of the participants reported motion sickness or nauseas when playing in the VR setting. Nonetheless, some of them reported fatigue on the arm used to play as a consequence of holding the hand in the air and constantly repeating the same gesture to shoot balls. For this reason, participants in the actual study were instructed to relax the arm when possible (i.e.: when letters were displayed on top of the doors). Finally, many participants highlighted the importance of the music and audio effects to achieve a good level of immersion.

4.4 Data Collection

This section describes the subjective (self-reported) and objective (physiological and motion) data collected during the present study, as well as the sensors used. Finally, the data normalisation and features extracted from physiological signals are explained.

4.4.1 Participants

Thirty participants, 15 male and 15 female, with mean age of 26.43 (SD: 4.8) were randomly assigned to one of the interaction modes in their first session. None of the selected participants had been diagnosed with any learning difficulty such as Dyslexia. 43% of the participants reported to have played video games between zero and two hours the week before the study, while 33% did not play any. All participants reported that they liked the game overall in both interaction modes, except three participants who disliked the game in their second session when playing in Desktop setting. The motivation of 60% of the participants when playing video games was just for fun and 33% reported to play "*to kill the time*". 63% of the participants had never used a HMD before or experienced any VR content. Hence, VR was a novel technology for most of the participants, which could have increased their motivation and the intensify of the affective states experienced, as reported in previous research [132].

Most of the participants (87%) found the Desktop interaction difficult to manage, and only 27% struggled in VR. This was mainly due to the different location of Leap Motion (LM) in Desktop and VR (see Fig. 4.2). The placement of this sensor in front of the HMD made the VR interaction easier and more natural to control, while the Desktop interaction was reported to be less natural and less comfortable.

4.4.2 Questionnaires

Three questionnaires were designed to collect subjective data from the participants (see Appendix B). At the beginning of the first session, participants completed a Pre-Experiment Questionnaire that gathered information about their background, gaming habits and prior gaming experience. This questionnaire also asked participants whether they had been diagnosed with any learning difficulty such as Dyslexia or Attention Deficit and Hyperactivity Disorder (ADHD).

After each difficulty level, participants completed a post-condition questionnaire and self-reported their levels of valence and arousal. This questionnaire contained the In-game version of the Gaming Experience Questionnaire (GEQ) as well as three questions about their levels of engagement, immersion and motivation. The In-game version of the GEQ is a concise version of the core GEQ, specifically designed to evaluate the game experience at multiple intervals during the game playing. This short version consists on 15 items that evaluate the 7 components presented in the core GEQ. These components are: Competence, Flow, Sensory and Imaginative Immersion, Tension, Challenge and Positive and Negative Affect. Valence and arousal were reported using the Affective Slider

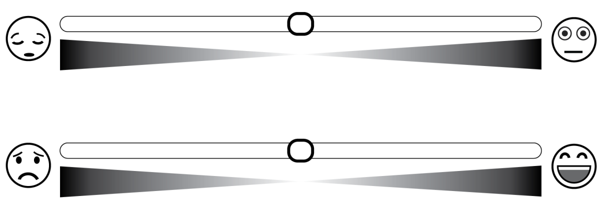

Figure 4.5: Affective Slider [21].

[21] (see Fig. 4.5), a modern digital version of the Self-Assessment Manikin [28].

Once all the levels were completed, participants completed a last post-session questionnaire, which included the Post-game module of the GEQ. This module consists of 17 items that assess 4 components: Positive Experience, Negative Experience, Tiredness and Returning to Reality. The post-session questionnaire also asked participants about their level of immersion, engagement, presence and motivation overall in the interaction mode played. Participants also reported whether they liked or not the game and if the interaction was difficult to manage. Finally, participants rated each levels in terms of difficulty, boredom, enjoyment, arousal and focus. Once the two sessions were completed, participants reported in which interaction mode they were more motivated to success, focused, aroused, immersed, challenged and which setting they enjoyed the most.

4.4.3 Physiological and Motion Sensors

Two physiological and two motion sensors were used in this study. The physiological sensors were a heart activity sensor and an electromyograph (EMG). Polar H7, a wearable chest strap sensor, recorded participant's heart activity. This device have been used in other studies that looked at stress awareness for children with ADHD in breath-controlled biofeedback games [150]. Data from the Polar H7 device was recorded on the iPhone app HRVLogger [5] via Bluetooth. This application allowed to add markers for each difficulty level. The HRVLogger app provided raw and pre-processed RR and HR data, sampled at 1Hz, as well as multiple time and frequency domain HRV features such as root mean square of successive differences (rMSSD), average of normal-to-normal intervals (AVNN), standard deviation of normal-to-normal intervals (SDNN), or Low and High Frequencies (LF and HF) and the ratio between them (LF/HF). The time window to calculate HRV features was set to 30 seconds, the minimum allowed. This means the HRV features are computed every 30s.

The EMG sensor MYO recorded the forearm's electrical activity of the hand used to interact with the game. MYO is built of 7 sensors that measure the electrical resistance of the forearm muscles. This data informed about participants' hand muscle activation (i.e.: how much pressure or force participants put on their hand) while playing the game. However, since the gestures were different in VR and Desktop modes and thus, different muscles were active, the data logged on each interaction mode could not be compared between them.

The motion sensors employed were the Head-Mounted Display (HMD), only used in the VR setting, and the hand-tracking device, Leap Motion, was used in both settings. The data logged from Leap Motion was the position and velocity of the hand controlling the game. The HMD's rotation axes were recorded to measure participant's head motion while playing in VR. The rotation was recorded in quaternions which is composed by four dimensions (X, Y, Z and W), although the W dimension was unfortunately not logged in.

4.4.4 Data pre-processing

Prior to the analysis, each participant's HR and HRV were normalised dividing the mean of each difficulty level by their baseline mean. This normalisation procedure scales the values so that all participants have a common mean of 1. No other pre-processing was needed on the HR or HRV features as they were computed by the HRVLogger app [5]. Different HRV features such as the average of normal-to-normal intervals (AVNN), root mean square of successive differences (rMSSD) and the low-high frequency ratio (LF/HF) were selected for the analysis. These features were chosen based on previous research linking HRV and visuospatial WM performance [153].

The hand's muscle activation data recorded by the EMG (MYO) did not have to be normalised as it was automatically calibrated when worn and synchronised with the computer. The mean level hand's muscle activation in each difficulty mode was calculated averaging the value each sensor by 7, the number of sensors measuring the muscle's activity.

WM capacity and WM performance scores were also normalised dividing the number of letters recalled correctly by the total number of letters presented, which results in the percentage of letters recalled correctly.

4.5 Analysis and results

This section analyses participants' player experience and affective states playing *Memory Break* in Desktop and VR. This analysis also explores the effects of arousal, valence and immersion on participants' WM performance. Since participants's WM capacity can affect their WM performance, this analysis examines individual differences in players' performance dividing participants in two groups depending on their WM capacity (low vs high) as measured with the AOST.

4.5.1 Statistics

The statistical analysis of this study was performed using RStudio 1.0.1. Firstly, Levene's tests were carried out to assess the normal distribution of the data. None of the tests were significant, which suggests the data is normally distributed. Before assuming the homogeneity of variance of the data, some linear models with the dependent variables of interest were created to plot and assess the distribution of the residuals. The Normal Q-Q plots did not show any violation of the homogeneity of variance, assuming the data is homoscedastic and thus, normally distributed.

To evaluate significant mean differences between interaction modes and between difficulty levels, repeated measures Analysis of Variances (ANOVA) were carried out. This test is useful to assess whether there is mean statistically significant differences in a certain variable across two or more groups or conditions. However, ANOVA tests only inform if statistical differences exist. They do not indicate in which conditions the differences are between. Further post-hoc repeated measures and independent sample t-tests were performed where relevant to evaluate how difficulty levels differed between them. Due to multiple comparisons, Bonferroni corrections [4] were applied where one variable was tested against more than one condition. The new significance level after the Bonferroni correction was set at $p < .017$.

Spearman's correlations (rho) were performed to assess monotonic relationship between two variables. Since most of the variables tested in this study were ordinal, this method was used to perform correlations between dependent and independent variables. Furthermore, Spearman's rank correlation does not make any assumption about the distribution of data, which makes it more robust.

Linear Mixed Effects (LME) models were created to perform regression analysis. These regression models are normally used when dealing with repeated measures variables. LME models are becoming popular recently due to their flexibility to account for fixed and random effects. Likelihood ratio test were performed to evaluate the goodness of fit of the models created. This ratio test indicate if the model proposed is significantly different from the null model and therefore, whether the parameters of the LME model contribute to predict the dependent variable. A full description of the LME models is reported in the results section as well as the Chi-square values, degrees of freedom and p-values resulting from the likelihood ratio test.

4.5.2 Interaction modes and difficulty levels

The analysis reported in this section compares participants' self-reported and physiological data while playing in Desktop and in VR. The only three participants that did not like the game, reported it after playing the Desktop setting in their the second session, which could be due to the difficulty controlling the

[4] The Bonferroni correction is used to avoid Type I errors when making multiple comparisons. It is calculated by dividing the significance level (in this case 0.05) by the number of tests that are being performed.

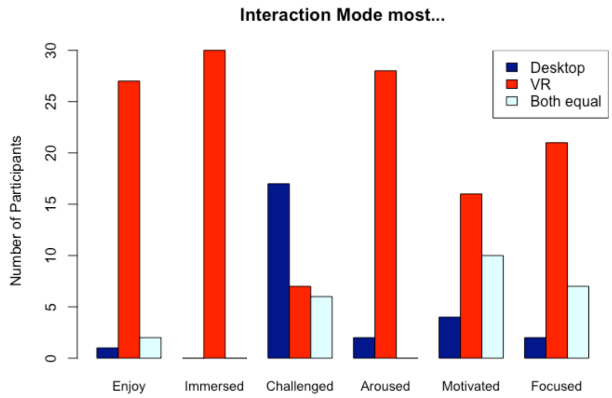

Figure 4.6: Post-Session Questionnaire: Comparison of self-reported variables between interaction modes.

interaction and a comparison with the VR setting.

Participants reported VR as the interaction they enjoyed the most, being also the setting where they felt more immersed, aroused and focused (Figure 4.6). VR was also the mode where most participants reported to be more motivated to play, while one third of the participants reported to be equally motivated in both interaction modes. As shown in Figure 4.7, most of the participants felt more bored in level 1 in both interaction modes, whilst the third level was the most difficult one and where they felt more aroused. Players reported they enjoyed similarly levels 2 and 3 on both modes. Desktop was reported challenging due to the difficulties in controlling the interaction, while only 27% reported to struggle controlling the VR interaction. Participants described the gesture interaction used in Desktop as uncomfortable and "*not very natural*". This indicates that participants experienced higher levels of challenge in Desktop due to the environment and interaction, not by the game itself.

Two-way repeated measures ANOVAs with Bonferroni corrections were carried out to test simultaneously for significant differences in each difficulty level. Engagement showed a significant difference between difficulty levels ($F= 7.25$, $p<.01$) and interaction modes ($F=18.91$, $p<.001$), while immersion ($F=25.41$, $p<.001$) and motivation ($F=16.95$, $p<.001$) were only significantly different between interaction modes, being both higher in VR.

The normalised mean HR was significantly higher in VR compared to Desktop ($F=9.70$, $p<.001$). In terms of difficulty levels, level 3 evoked the highest HR ($F=24.53$, $p<.001$). Nevertheless, these results could be affected by how much participants had to move their hand and head (in VR) to succeed in each difficulty level. Among the HRV features extracted, only AVNN showed sig-

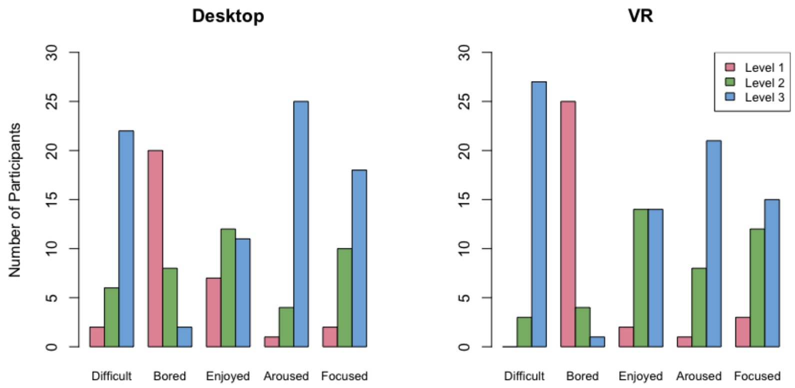

Figure 4.7: Post-Session Questionnaire: Comparison of self-reported variables in each difficulty level in both interaction modes.

nificant differences between difficulty levels (F=35.42, p<.001) and interaction modes (F=8.76, p<.01), while rMSSD did only for difficulty levels (F=8.96, p<.001). The self-reported levels of arousal (F=12.73, p<.001) and valence (F=19.70, p<.001) were statistically lower in Desktop than VR. Furthermore, arousal (rho=0.48, p<.001) and valence (rho=0.56, p<.001) correlated significantly with the reported immersion in both settings, indicating that high levels of immersion lead to an increase in self-reported valence and arousal. These results are in line with previous work measuring affective states in VR environments [132].

The EMG data was analysed separately in each interaction mode as the gestures were not the same in Desktop and VR and therefore different muscles were activated. Significant differences were found between difficulty levels in Desktop (F=31.18, p<.001) and VR (F=66.01, p<.001). As shown in Figure 4.8 , the highest muscle activation was observed in the hard level of both interaction modes. However, this could be explained by the number of shooting gestures participants had to do in the hardest level as there were more obstacles.

4.5.3 Working Memory performance

An important aspect of this study looked at differences in the normalised WM performances. A first ANOVA analysis on each interaction separately showed a higher but not significant difference in VR (F=2.20, p=.09) than in Desktop (F=0.52, p=.67). The greater and almost significant differences in VR encouraged a further analysis to assess for statistical differences in the difficulty levels only in this interaction mode. Post-hoc repeated measures t-tests between the

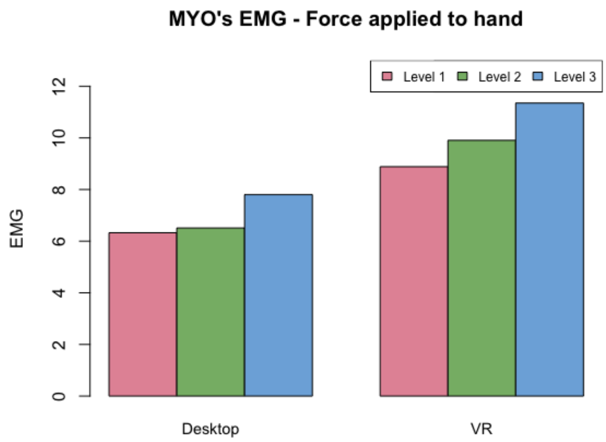

Figure 4.8: Hand's muscle activation (i.e.: pressure exerted in the hand) in each difficulty level of Desktop and VR
NOTE: According to MYO's documentation, EMG data is provided in unitless format called "*activation*"[1], although EMG is normally measured in µVolts

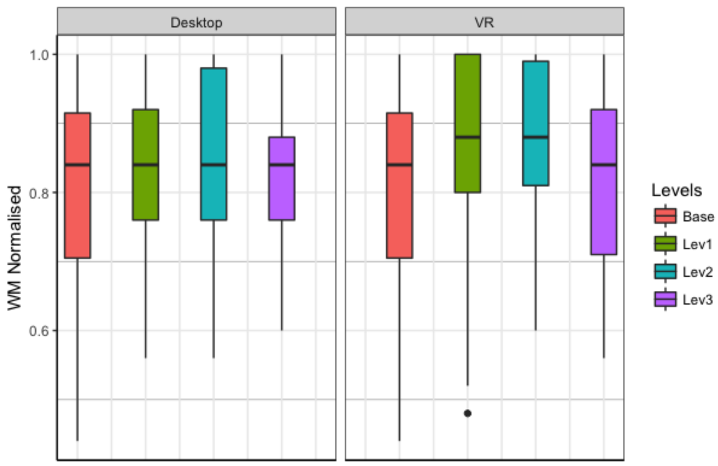

Figure 4.9: WM performance of all participants in each difficulty level

WM capacity and WM performance scores were performed. Due to the multiple comparisons between the WM capacity baseline and WM performance in each difficulty level, Bonferroni corrections were made, setting a new significant level at *p< .017*. Significant differences were found in easy (t=-2.18, p<.05) and medium (t=-2.15, p<.017) difficulty levels (see Fig. 4.9), although only the later is significant when applying Bonferroni corrections on the significance level. No significant differences were found in the most difficult level (t=-0.80, p=.43). Moreover, the normalised WM performance scores presented a weak but significant positive correlation with the self-reported competence (rho=0.23, p<.01), and a negative correlation with tension (rho=-0.28, p<.001).

Based on the normalised WM capacity baseline, as measured with the AOST [56] test at the beginning of the study, participants were divided into two groups: low and high WM. Those with a normalised WM baseline lower than the overall median (0.83) were assigned to the low WM group, and those above the median to the high WM group. This resulted in 13 subjects in the low WM group and 17 in the high WM. A repeated measures ANOVA was performed for each group, only showing significant differences in the low WM group for the difficulty levels (F=4.96, p<.01). A post-hoc paired t-test analysis assessed statistical differences for the difficulty levels in the low WM group compared to their WM capacity baseline. Participants with low WM achieved significantly better WM scores in VR (Fig. 4.10) in the easy (t=-2.18, p<.05) and medium (t=-3.22, p<.05) difficulty levels. Significant differences using two-samples t-tests were also found in the HRV features extracted between these groups. Similar to previous research reported in a visuo-spatial WM task [153], the LF/HF ratio (t=4.31, p<.001) was significantly higher for the low WM group.

4.5.4 Effects of valence and arousal on WM

This section investigates Figures 4.10 and 4.11, analysing the self-reported levels of valence and arousal of the low and high WM groups. As shown in Figure 4.11, arousal levels increased for all participants in all difficulty levels. According to Bennion et al. [19], arousal has beneficial effects on WM up to a certain point, after which it has a negative effect. Looking at the self-reported arousal of the two groups, the highest level of arousal is observed in the third and most difficult level of both interaction modes, which correspond to the lowest WM scores, specially for the high WM group in Desktop setting.

The levels of valence showed more interesting results, correlating significantly with WM performance (rho=0.19, p<.01), and being particularly pronounced for the high WM group (rho=0.39, p<.001). This indicates that high levels of positive valence improved WM performance. It was observed that when

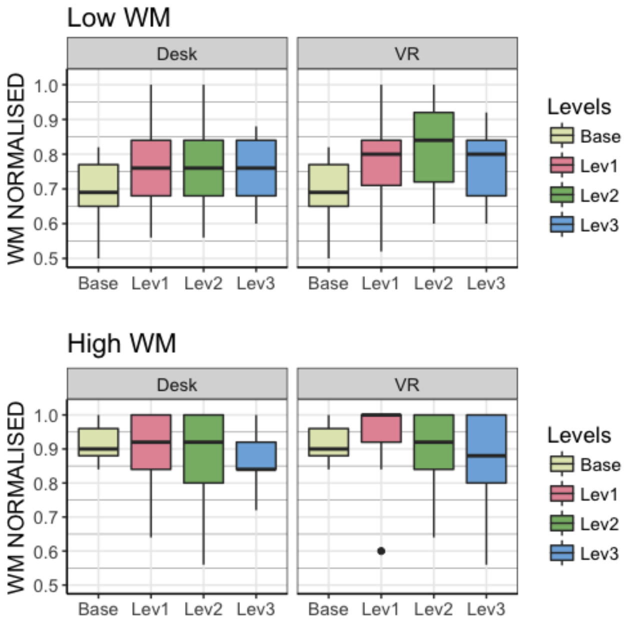

Figure 4.10: WM performance of Low and High WM groups in each difficulty level

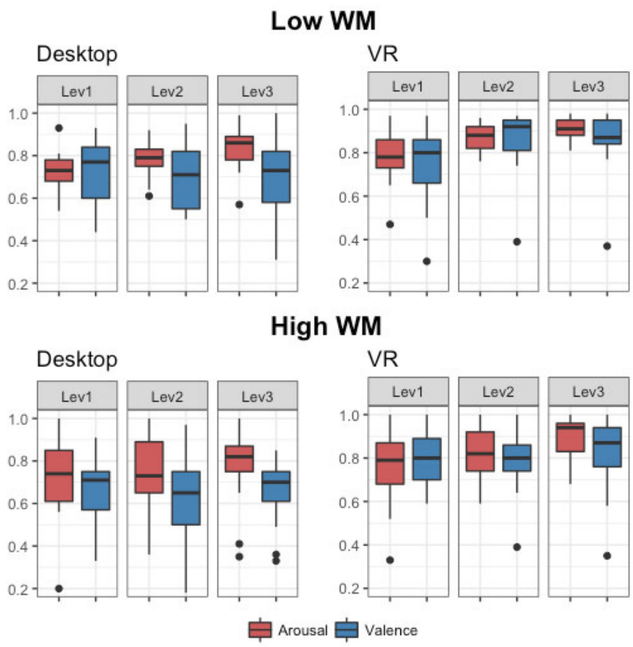

Figure 4.11: Self-reported arousal and valence of Low and High WM groups in each difficulty level

valence and arousal are both high, i.e. when participants were challenged but feeling successful, they obtained their best WM score. More specifically, WM performance can improve when the player is in a state of enjoyment or flow, described by Csikszentmihályi as the optimal experience [45]. However, the performance of the low WM group in Desktop's level 1 is explained by the order of levels played, since half of the low WM participants played level 1 last, having had more experience playing the game and being more relaxed.

Linear Mixed Effects regression models were used to predict WM performance and the immersion of participants. We developed a model using the self-reported valence and arousal as fixed effects and subjects as random effects to predict WM performance. Since we are interested in the relationship between WM and affective states, arousal and valence were the only predictors for the WM model. As seen in Table 4.1, valence and arousal were significant predictors of WM performance. However, the estimate of arousal is negative while valence is positive. This means that high levels of arousal may have negative effects on the WM performance, whilst high levels of positive valence can have positive effects. A likelihood ratio test was performed, showing the goodness of fit of this model (Chi-sq=27.17, p<.001).

Using the same random effects, the model to predict immersion used WM, the interaction mode and HRV rMSSD as fixed effects. The inputs of this model explored the relationship between WM, HRV and immersion. The interaction mode and HRV rMSSD were significant predictors of immersion, but not WM, although it improved the model (see Table 4.2). The negative estimate of HRV rMSSD suggests that high levels of HRV may have negative effects in the immersion, which means that participants being more relaxed, tend to be less immersed. The likelihood ratio test was also significant for this model (Chi-sq=46.92, p<.001).

Table 4.1: LME model of WM

	Estimate	**Std. Error**	**df**	**t**	**p**
(Intercept)	19.54	1.17	175	16.71	<.001
Valence	6.13	1.12	161	5.47	<.001
Arosual	-3.79	1.52	170	-2.49	<.01

Table 4.2: LME model of immersion

	Estimate	**Std. Error**	**df**	**t**	**p**
(Intercept)	1.98	.43	170	4.55	<.001
WM	.03	.02	179	1.55	.12
IM (VR)	.62	.09	151	6.66	<.001
HRV rMSSD	-3.79	.19	179	1.88	.06

4.6 Discussion

4.6.1 WM in VR

The results presented evidence of how game playing in VR can improve WM performance and intensify the self-reported affective states of the players. While being immersed and engaged in a VR game, players make better use of their cognitive resources as they are more activated and motivated. This effect was stronger on those with a lower WM capacity, who showed a significant improvement in their WM performance when playing the easy and medium difficulty levels in VR.

Some participants reported that the different sessions' structure affected their WM performance. As the first session was longer due the AOST test preceding the game, some participants informed this affected their cognitive skills. Notwithstanding, there was disagreement between participants regarding the effects of the AOST on their WM performance in *Memory Break*. Whilst some participants manifested that the AOST caused them cognitive fatigue, others reported it activated them, being more ready for the WM tasks presented in the game. Practice effects of participants playing *Memory Break* again in their second session could have also impacted on the results as they already had experience playing the game. Nevertheless, the randomised controlled trial design of this study should have removed these effects.

It is important highlighting the differences in the AOST, which measured participant's WM capacity, and the WM task in *Memory Break* could. Some participants reported that the AOST test was more challenging than the WM task in *Memory Break* as the first one involves math problems and letters, which share the verbal cognitive resources. This could have a positive impact on the results as those with a low WM would perform better in the WM task within *Memory Break*.

4.6.2 Arousal, valence and WM

The high levels of arousal and valence reported in VR had a positive effect on the player's cognitive performance, possibly enhancing the capture and encoding of information. However, high levels of arousal had negative effects on the player's WM when not accompanied by high levels of positive valence. In other words, when players were highly aroused but not enjoying the experience, the WM performance decays. Likewise, if players are experiencing high levels of arousal and positive valence, player's WM performance may improve.

These results can be linked to the flow theory proposed in [45] known as a state of full immersion and engagement triggered when a subject's skills can overcome challenges, creating a positive experience and positive affect. When players are in this flow state, highly focused and enjoying the game, they make an optimal use of their cognitive skills and thus, a better WM performance is achieved. Hence, the game has to be challenging enough but not too difficult in order to keep the player engaged, motivated and focused on the given task.

Significant differences were also found on the physiological signals measured. The normalised mean HR was significantly higher in VR. The normalised mean HRV features AVNN and rMSSD also showed significant differences between interaction modes, being lower in VR than Desktop. AVNN was also significantly different between difficulty levels. The hand's muscle activation, measured with the EMG, was significantly higher in the hard level on both interaction modes. These results are in line with previous findings suggesting players had a higher muscle activation (i.e.: pressure on the hand when the difficulty increase and felt more challenged. Nonetheless, these results might be affected by the number of shooting gestures needed to succeed in the hard level. Significant differences found in the normalised mean HR and HRV could also have been affected by the greater amount of movement required in the most difficult level.

4.6.3 Implications

Game playing in VR can improve the player's WM performance. Higher level of self-reported immersion, engagement and motivation when playing in VR have a positive effect on the player's cognitive skills. This effect is especially pronounced on those subjects with a low WM capacity. The results of this study suggest that valence and arousal can have positive effects on the WM performance. High levels of arousal accompanied by high levels of positive valence help players to make a better use of their cognitive resources and therefore improve their attention and WM performance. However, when high levels of arousal are not accompanied by positive valence, creating a negative experience, players have a worse WM performance. This is related to the theory of flow [45], which argues that there should be a balance between the challenges given and the player's skills to foster an optimal affective state of enjoyment that sustains engagement.

An important limitation of this study was the difference in placement of the hand motion tracking device that led to different gestures in Desktop and VR settings. The reported difficulties in controlling the Desktop interaction might have affected the level of immersion, even though the majority of the participants liked the game in this setting. It is also important to mention that different results could have been obtained using other types of games requiring higher verbal or spatial cognitive demands.

4.7 Summary

This study has investigated the effects of game playing in Desktop and VR settings on WM performance, as well as the effects of the self-reported arousal and valence on WM performance. A custom video game with a hand-gesture based interaction was developed for both Desktop and VR. Three difficulty levels were created to induce different levels of arousal, maintaining the same memory load for all levels. Physiological and self-reported measures of valence and arousal among other variables such as immersion and engagement were collected.

Higher levels of self-reported immersion while playing *Memory Break* in VR had a positive effect on the player's WM performance. This improvement was particularly pronounced in participants with low WM capacity, as measured with the AOST test [56]. Suggestions have been presented of how self-reported affective states can be beneficial for WM when playing a video game. High levels of arousal and positive valence can create a positive experience, leading players to a flow state [45] that may have a positive impact on the player's WM performance.

This research proposes to work towards a closed-loop video game in VR that includes the player's affective states in the adaptive loop in order to improve the adaptation. The ideal affective video game for cognitive training should keep the player in an optimal affective state while challenging his or her cognitive skills. Since VR is known to increase the level of immersion [132], a VR game can potentially help players to achieve a better WM performance. The next chapter focuses on implementing a machine learning algorithm in the VR game *Memory Break* to detect the player's arousal and valence in real-time, in order to automatically adapt the difficulty level to improve the player's WM performance.

Chapter 5

The Adaptation Engine

This chapter documents the design, implementation and testing of the affect recognition system as well as of the adaptation engine. Whilst the affect recognition system detects the player's affective states, the adaptation engine uses the system's output and the player's performance to adapt the game's difficulty in real-time. The adaptation engine, described in detail in section 5.1, aims to sustain the player in an optimal affective state in order to improve the WM performance. The system analyses various motion and physiological features extracted from different sensors to recognise the player's arousal and valence levels. Given the results reported in the previous chapter, the remainder of this research is focused on VR, disregarding the Desktop interaction mode. The implementation of the adaptation engine in *Memory Break* is described in section 5.2, followed by a series of three pilot studies that tested it, reported in section 5.3. Finally, the decision rules for difficulty adaptation are outlined in section 5.4.

5.1 Affect recognition system design

Using the data collected (see Figs 5.1 and 5.2) in the study reported in the previous chapter (Chapter 4), two machine learning algorithms were trained for arousal and valence classification. Table 5.1 summarises the sensors used in the previous study as well as the type, sampling frequency, body part sensed and data collected from each sensor.

5.1.1 Data pre-processing

The data collected was cleaned and pre-processed in MATLAB prior to the feature extraction. First, the data collected from the HMD was down-sampled

Table 5.1: Sensors and data collection.

Sensor	**Type**	**Sampling Frequency**	**Body Part**	**Data Collected**
Leap Motion	Motion	50Hz	Hand	Velocity
HMD	Motion	75Hz	Head	Rotation
Polar H7	Physiological	1Hz & 0.03Hz	Heart	HR and HRV
MYO	Physiological	50Hz	Forearm Muscles	Muscle activation

Note: HR and HRV were sampled at different frequencies.

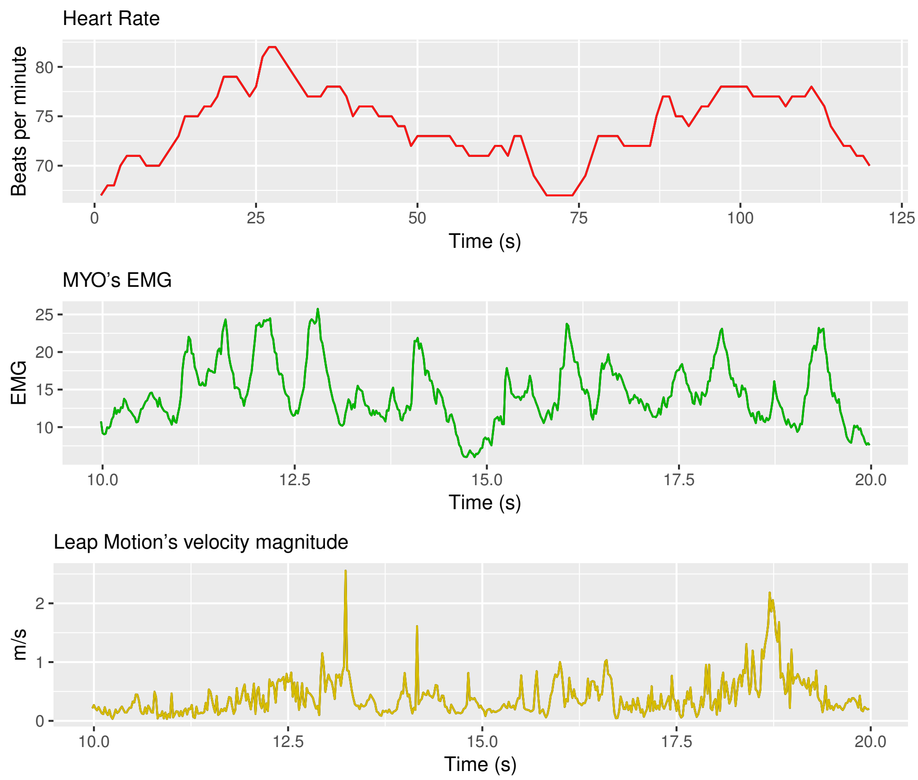

Figure 5.1: Sample data of HR (top), MYO's EMG muscle activation (middle) and Leap Motion's velocity magnitude (bottom).
NOTE: According to MYO's documentation, EMG data is provided in unitless format called "*activation*"[1], although EMG is normally measured in µVolts

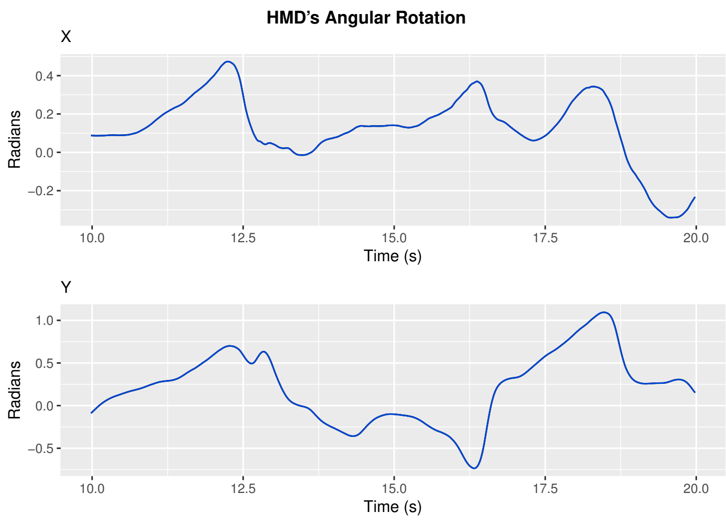

Figure 5.2: Sample data of HMD's angular rotation in X and Y axis.

to 50Hz using a linear interpolation between the values. This put the HMD and Leap Motion's (LM) data at the same sampling rate. One participant was removed from the dataset as he reported high levels of arousal but his normalised HR during play was below his resting HR. Even though the participant was equally or more relaxed playing than resting, his reported levels of arousal were always greater than 0.87 (the overall median of arousal). Since 3 data points were gathered for each participant in Study 2, one per difficulty level, the original data set of 90 data points was reduced to 87.

The HMD's angular rotation (see Fig. 5.3) data was recorded in Unity in quaternions instead of Euler angles or radians. Quaternions are a four-dimensional (X, Y, Z and W) number system that extends complex numbers to

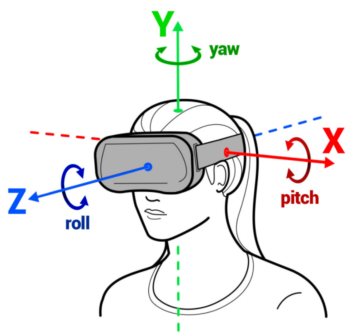

Figure 5.3: HMD's rotation on X (pitch), Y (yaw) and Z (roll) axes

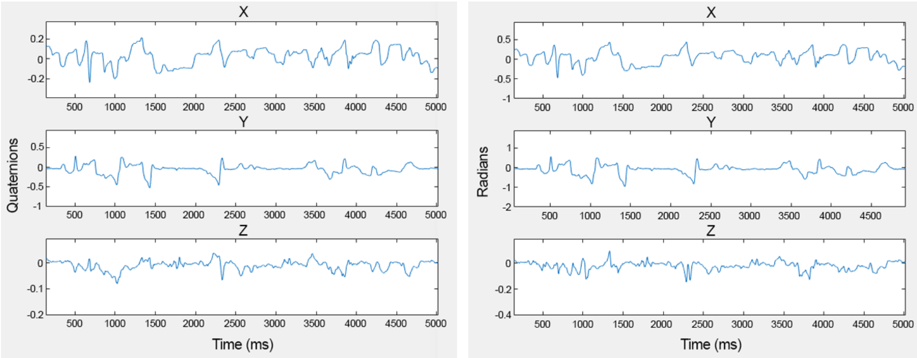

Figure 5.4: Conversion of HMD's rotation in quaternions (left) to radians (right). From top to bottom, the figure shows the rotation on the X, Y and Z axes. Note how the conversion from quaternions to radians did not recompose the Z axis satisfactorily.

describe that rotation in a three-dimensional space that prevents gimbal lock[1], which is the loss of one degree of freedom in a three-dimensional rotation. The HMD's data was converted to radians to calculate the angular velocity and acceleration (see Fig. 5.4). However, one of the quaternion's dimension (W) was missing when recording the data. Replacing the W dimension with a vector of ones, the X and Y axes were recomposed with an acceptable error margin, but not the Z axis. As shown in Figure 5.4, this introduced some noise in the Z axis when converting quaternions into radians so it was disregarded. Although removing the Z axis rotation (roll) implies a reduction of accuracy on the recomposition of head movements, researchers have only used rotations on X and Y axes to detect affective states from head motion [17]. Therefore, the HMD's angular velocity and acceleration magnitudes were calculated with the X and Y axes only.

Once the HMD's data was converted to Euler angles, the data collected from LM, HMD and MYO was smoothed using a moving average filter with a window of 5 samples. This filter preserved the envelope of the signal, which was used to detect peaks and valleys. No filtering was applied to either HR or HRV features as they were already pre-processed and smoothed. Finally, since there was no interest in the direction of the head or hand motion, absolute values were computed for LM's velocity and HMD's angular rotation.

5.1.2 Feature extraction

Various features were computed from the motion (HMD and LM) and physiological (HR and EMG) sensors used in the previous study. Table 5.2 details

[1]https://en.wikipedia.org/wiki/Gimbal_lock

Table 5.2: Features extracted of each sensor.

Sensor	**Activity Measured**	**Features Extracted**
HR and EMG	Cardiac Activity and Hand's Muscle Activation	Mean
		Standard deviation
		Maximum
		Mean of 1st derivative of raw signal
		Mean of 1st derivative of normalised signal
		Mean of 2nd derivative of raw signal
		Mean of 2nd derivative of normalised signal
HR	Cardiac Activity	Mean of HRV AVNN
		Mean of HRV rMSSD
		Mean of HRV LF/HF Ratio
HMD and LM	Head and Hand's Velocity and Acceleration	Mean
		Standard deviation
		Maximum
		Mean of peaks' width as well as of the initial and final slopes
		Standard deviation of peaks' width as well as of the initial and final slopes
		Maximum of peaks' width as well as of the initial and final slopes
		Number of peaks
LM	Hand's motion	Number of zeros[1]

[1] Time the playing hand is relaxing, out of Leap Motion's field of view.
Total number of features of each sensor: HMD: 26; LM: 27; EMG: 7; HR: 10.

all features extracted. This section describes the features extracted from each sensor.

Motion features

Motion features were calculated using the velocity and acceleration magnitudes of the HMD and the hand used to interact with the game, tracked with LM. Only the hand's palm motion was recorded for feature extraction, no fingers were analysed. Prior to the extraction of motion features, a z-score normalisation was applied to the motion's velocity and acceleration of both HMD and LM, having zero mean and one standard deviation (Formula 5.1). Since velocity has a magnitude and direction, absolute values were calculated for each axis of both the HMD and the hand's motion, disregarding the direction. Acceleration is the rate of change of velocity of an object with respect to time [2]. Magnitude and direction were preserved as they provided information about the acceleration and deceleration. Once velocity and acceleration were calculated, the axial components of each feature were summed and square-rooted to calculate the magnitude of both features. The acceleration's magnitude, also called quantity of motion (QoM), has been one of the most successful motion features to classify emotions [35]. Mean, maximum and standard deviation of both velocity and acceleration were computed for the HMD and hand's motion.

When playing in VR, participants had to keep their hand suspended in front of the HMD so Leap Motion could track their hand gestures. In order to avoid arm fatigue, participants were encouraged to relax and put the arm down when possible, disrupting the hand's tracking. This absence of hand detection turned into another feature describing how much time participants had their hand down. This feature was computed counting the number of zeros (i.e. no hand detected) in Leap Motion's data.

$$\overline{x} = \frac{x_i - \mu}{\sigma} \tag{5.1}$$

Following Castellano's mathematical model to analyse gestural expressivity [35], various motion features were derived such as the peaks' slopes and duration of the velocity and acceleration of both the HMD and Leap Motion (LM). Since there was no specific gestures labeled with emotions, as in Castellano's work, a peak detection algorithm was used to find all the peaks with a threshold of 2 and a minimum distance between peaks of 0.5. Prior to the extraction of these features, velocity and acceleration were normalised (Formula 5.1), having zero mean and one standard deviation. To calculate the initial and final slopes as well as the duration of the peak, it was necessary to detect the valleys before

[2]https://en.wikipedia.org/wiki/Acceleration

and after each peak found. This was accomplished by inverting the signal, so that what used to be valleys would present as peaks. Another peak detection algorithm was applied to the inverted signal, with a threshold of -0.1 and no minimum peak distance. This returned all the valleys in the original signal smaller than 0.1. For each peak's location, the valleys before and after were identified according to the valley's location. Then the initial slope was calculated subtracting the value of the valley before the peak from the peak's value, and then dividing by the time difference between this valley and the peak. The same method was used for the end slope calculation. Finally, the peak's duration - in milliseconds - was calculated by subtracting the timestamp of the valley before the peak from the timestamp of the valley after it. Mean, standard deviation and maximum values were calculated for the peaks' duration, start and final slopes. The number of peaks detected were also used as a feature. This resulted in a total of 36 features for the HMD, and 37 for the hand's motion. A z-score normalization was applied to all computed features using their own mean and standard deviation.

Physiological features

Before computing the physiological features, participants' HR and HRV were normalised using their own baseline recorded while resting at the beginning of each session. No normalisation was needed for the EMG as it was calibrated in every session.

Six features were computed for the physiological sensors (HR and EMG) following Picard, Vyzas and Healey's proposed features [122] to measure emotions from physiological signals (see Formulas 5.2-5.7). The main advantage of these features is that they can be easily computed in real-time, which makes them suitable for implementation in *Memory Break*. The maximum of each signal was also added to the feature sets.

The HR feature set included the mean of three heart rate variability (HRV) features: the root mean squared of successive differences (rMSSD), the average N-N intervals (AVNN) and the low to high frequency ratio (LF/HF). Since the HRV recordings were made every 30s (sampling rate of 0.03Hz), the same time as one game's section, only one HRV observation could be considered for each section. Therefore, no further features could be obtained from only one HRV observation. The raw values of AVNN, rMSSD and LF/HF ratio were included in the set of HR features. This resulted in a total of 10 HR features and 7 EMG features. Again, a z-score normalization (Formula 5.1) was applied to all physiological features using their own mean and standard deviation.

$$\mu = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{5.2}$$

$$\sigma_x = (\frac{1}{N-1}\sum_{n=1}^{N}(X_n - \mu_x)^2)^{1/2} \tag{5.3}$$

$$\delta_x = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|X_{\mathrm{n}+1} - X_{\mathrm{n}}\right| \tag{5.4}$$

$$\tilde{\delta}_x = \frac{1}{N-1}\sum_{n=1}^{N-1}\left|\tilde{X}_{\mathrm{n}+1} - \tilde{X}_{\mathrm{n}}\right| = \frac{\delta_x}{\sigma_x} \tag{5.5}$$

$$\gamma_x = \frac{1}{N-2}\sum_{n=1}^{N-2}\left|X_{\mathrm{n}+2} - X_{\mathrm{n}}\right| \tag{5.6}$$

$$\tilde{\gamma}_x = \frac{1}{N-2}\sum_{n=1}^{N-2}\left|\tilde{X}_{\mathrm{n}+2} - \tilde{X}_{\mathrm{n}}\right| = \frac{\gamma_x}{\sigma_x} \tag{5.7}$$

5.1.3 Labelling and classification

The aim of the machine learning for affect recognition was to be able to recognise at least four distinct affective states. Due to the limited amount of data available from the previous study (87 data points), it was split into two classes for arousal (high vs low) and valence (positive vs negative) separately. Since the distribution of the self-reported affective states was considerably skewed (see Fig. 5.5), the median of the reported valence (0.85) and arousal (0.87) was used as a threshold to divide arousal and valence into two classes. Forty-nine data points were categorised as low arousal and thirty-eight as high arousal, and forty-nine data points for negative valence and thirty-eight for positive valence. Following Russell's model of affect [138] (see Fig. 2.1), the combination of these arousal and valence classes result in 4 emotions: low arousal negative valence (bored), low arousal positive valence (relaxed), high arousal negative valence (frustrated) and high arousal positive valence (excited). These 4 emotions, the two classes of valence and the two classes of arousal were used to label the data for classification. Table 5.3 show the number of data points for each classification label.

Eight different machine learning algorithms were tested for arousal and valence classification using Weka [169]. These algorithms were: 1) probabilistic (Naïve Bayes); 2) linear (Logistic Regression); 3) non-linear models (Support Vector Machines (SVM) with RBF kernel); 4) Neural Networks (multilayer perceptron); 5) lazy learning (k-nearest neighbour - KNN); 6) meta-classifiers (AdaBoost with Naïve Bayes as base classifier); 7) decision trees (J48); and 8) ensemble decision trees (Random Forests) [94]. These algorithms were tested with their default parameters and a 10-fold cross-validation. All features computed were normalised in MATLAB (Formula 5.1) prior to these tests, having a zero mean and a standard deviation of one.

5.1.4 Section selection

As described in section 4.3, each game play was divided into five sections consisting of 30s of game play and a WM trial. In the previous study, players self-reported their arousal and valence levels at the end of each game play and not during each section as in the present study. Since the aim is to infer the player's affective states at the end of each section to adapt the next section's difficulty, the section that best represents the self-reported affective states had to be selected. This is necessary since the machine learning algorithms should be trained with data collected over the same amount of time that it will be tested against. The data collected during the WM trials (ie: between doors) of

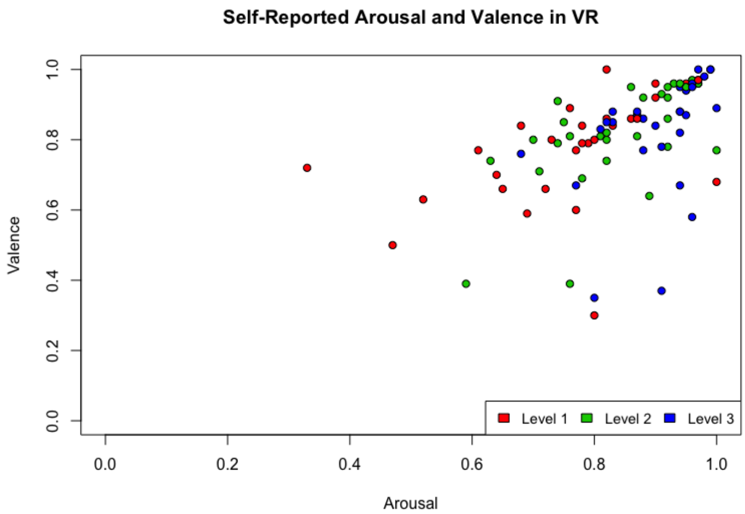

Figure 5.5: Self-reported arousal and valence in VR in Study 2

Table 5.3: Arousal and Valence labels distribution

Arousal	Valence	Emotion Label	Data Points	Percentage
Low	Negative	Bored	38	43.68%
Low	Positive	Relaxed	11	12.64%
High	Negative	Frustrated	11	12.64%
High	Positive	Excited	27	31.04%

Table 5.4: Classification accuracies (in %) of all machine learning algorithms tested in each section of *Memory Break* in Study 2

	Section 1			Section 2			Section 3			Section 4			Section 5		
	4 Emotions	**Arousal**	**Valence**	**4 Emotions**	**Arousal**	**Valence**	**4 Emotions**	**Arousal**	**Valence**	**4 Emotions**	**Arousal**	**Valence**	**4 Emotions**	**Arousal**	**Valence**
Naïve Bayes	35.63	59.77	62.06	37.93	66.66	57.47	47.12	**67.81**	65.51	33.33	57.47	54.02	33.33	62.06	52.87
Logistic Reg.	33.33	55.17	52.87	25.29	52.87	48.27	31.03	55.17	54.02	35.63	56.32	5862	37.93	54.02	48.27
SVM	**52.87**	64.36	66.66	51.72	62.07	**73.56**	**52.87**	64.36	58.62	45.97	56.32	54.02	45.91	55.17	52.87
NN	45.97	60.91	51.72	35.63	52.87	63.21	35.63	52.87	50.57	35.63	48.27	54.02	33.33	52.87	50.57
KNN	50.57	64.36	66.66	35.63	52.87	68.96	45.97	58.62	57.47	36.78	50.57	59.77	36.78	52.87	49.42
AdaBoost	35.63	59.71	62.06	39.08	62.06	54.02	42.52	56.32	54.02	37.93	49.42	55.17	34.448	59.77	50.57
Rand. Forests	51.72	66.66	58.62	47.13	51.72	66.66	50.57	64.36	57.47	44.82	60.92	63.21	44.82	58.62	49.42
J48	40.22	50.57	55.17	40.23	47.12	72.41	34.48	56.32	49.42	36.78	50.57	48.27	33.33	51.72	45.97
Mean	43.24	60.19	59.48	39.08	56.03	63.07	42.52	59.48	55.89	38.36	53.73	55.89	37.50	55.89	50

each section was ignored since our aim is to detect the player's affective states only looking at the time playing. This decision was motivated to generalise the affective states recognition to any VR game with a similar settings. Thereby, five sets of features (one per section) were created.

Using all the 70 physiological and motion features extracted, the aforementioned algorithms were tested with each of the five sections that comprise one game play. This subject-independent analysis was carried out using a ten fold cross-validation in Weka [169]. Furthermore, these tests were performed with the 4 emotions labels as well as for the 2 levels of valence and 2 levels of arousal separately (see Table 5.3). The results in Table 5.4 show that the best accuracy results were obtained in sections 2 and 3 with the arousal and valence labels. Even though section 3 showed a slightly better accuracy for arousal with Naïve Bayes (67,81%) than section 2 (66,66%), section 2 was selected as it showed the best accuracy for valence (73,56%) classification with SVM as opposed to Naïve Bayes (57,47%).

5.1.5 Feature and model selection

The contribution of each sensor (HMD, LM, HR and EMG) for arousal, valence and the 4 emotions classification with the aforementioned machine learning algorithms was investigated. Fourteen different models were created fusing the sensors' features in all possible combinations. (see Appendix C) This analysis is intended to select the sensors' features that best represent the self-reported arousal and valence levels.

The HR+EMG+HMD model had the best performance using SVM with the 4 emotions, achieving a 56.32% accuracy. However, the confusion matrix of this classification (see Table 5.5) shows that only two emotions, bored and excited, were successfully classified. These results seem reasonable as these emotions represent two completely different affective states. One possible explanation for these results is the limited amount of data available. Since only 2 out of the 4 emotions were classified, the 4 emotions labels were disregarded. Thus, the affect recognition system would make a decision-level fusion combining the arousal and valence classification outputs to detect the 4 emotions in order to make an adaptation decision.

On the other hand, only three models had the same or higher accuracy than using the full feature set for arousal classification. Whilst the HMD+HR model (36 features) achieved a 66,66% accuracy with Naïve Bayes and AdaBoost classifiers, the HMD+HR+LM (63 features) and HMD+EMG+LM (60 features) models provided an accuracy of 67,81% with Naïve Bayes and KNN. The best accuracy for valence classification (73,56%) was achieved by two models: the

HMD+EMG (33 features) with J48 and HMD+EMG+HR (43 features) with SVM. Since none of the models achieved significantly better accuracies than the Full model with all features, four models for arousal (Full model, HMD+HR, HMD+HR+LM and HMD+EMG+LM) and three for valence classification (Full Model, HMD+EMG and HMD+EMG+HR) were selected. This indicates the importance of fusing multimodal features (i.e: motion and physiological) to maximise the performance of the affect recognition system.

Feature selection was carried out in Weka using these models. Two selection methods were used: correlation selecting the top 10 ranked features, and the CfsSubsetEval algorithm, which '*evaluates a subset of features by considering the individual predictive ability of each one along with the degree of redundancy between them*' [63]. The number of features selected by the CfsSubsetEval algorithm depends on the set of features provided. The Wrapper algorithm for feature selection used by Castellano et al. [35] was discarded due to the extensive time it takes to complete for some machine learning algorithms such as Neural Networks (i.e., Multilayer Perceptron).

None of the feature selection methods improved arousal accuracy with the models selected. In order to select the best models for arousal classification, the average accuracy of the 8 algorithms tested was calculated for each model. While the HMD+HR+LM and the Full models achieved an average accuracy of 55% and 56% respectively, the HMD+HR and HMD+EMG+LM models achieved 59% and 58% average accuracy. Only the HMD+EMG model improved the valence classification accuracy to 77,01% when selecting the top 10 features using J48, followed by 75,86% accuracy using Random Forests or KNN. CfsSubsetEval also achieved similar accuracy selecting 5 features of the same model with SVM. Thus, two feature subsets of this model (HMD+EMG) with 5 and 10 features for valence classification, and two models (HMD+HR and HMD+EMG+LM) with all features for arousal classification were selected.

Table 5.5: Confusion matrix of SVM with the HR+EMG+HMD model (43 features).

	Bored	**Relaxed**	**Frustrated**	**Excited**
Bored	35	0	0	3
Relaxed	9	0	0	2
Frustrated	9	0	0	2
Excited	13	0	0	14

5.1.6 Model optimisation

The next step was to select and tune the classification algorithms to obtain the best performance. A parameter optimisation on all algorithms tested was performed, since some of them (i.e.: SVM) require parameter tuning to achieve their best performance [87]. The meta-classifiers CVParameterSelection and GridSearch were used in Weka to optimise parameters simultaneously. For example, the GridSearch was used to explore different values of Cost (C) and Gamma (G) parameters of SVM. Using the full dataset collected in the previous study (Chapter 4), parameters were manually fine tuned when the meta-classifiers improved the classification accuracy.

Both feature subsets of the HMD+EMG model improved the valence classification accuracy when optimising the parameters of certain algorithms. SVM achieved the best accuracy (81,60%) for valence classification with the 5 features subset of the HMD+EMG model. On the other hand, only the HMD+HR model significantly increased its accuracy from 59,77% to 74,71% with SVM. Thus, SVM was used for both arousal and valence classification. A decision fusion was applied after classification, combining the output of both SVMs in order to estimate the player's affective state.

5.1.7 Subject-independent analysis

Before testing the models in real-time, a subject-independent analysis was performed to check how the affect recognition system would perform with unseen data. A leave-one-subject-out approach was used, training the SVM algorithms with data from all except one participant (84 data points) and tested it with the excluded participant (3 data points). This procedure was repeated 29 times, one for each participant of Study 2. As expected, the average accuracy of the 29 runs decreased for both valence and arousal classification. Whilst valence classification decreased from 81,60% to 80,45% accuracy, arousal had a more significant accuracy decline from 74,71% to 67,81%.

5.2 Implementation

The implementation of the affect recognition system in *Memory Break* was realised with Accord [151], a .NET framework for machine learning embedded in Unity, the software used to develop the game. Two SVM algorithms with Radial Basis Function (RBF) kernels were trained for arousal and valence classifications, using the HMD+HR (36 features) and HMD+EMG (5 features) models selected. These algorithms were tuned using the parameters obtained

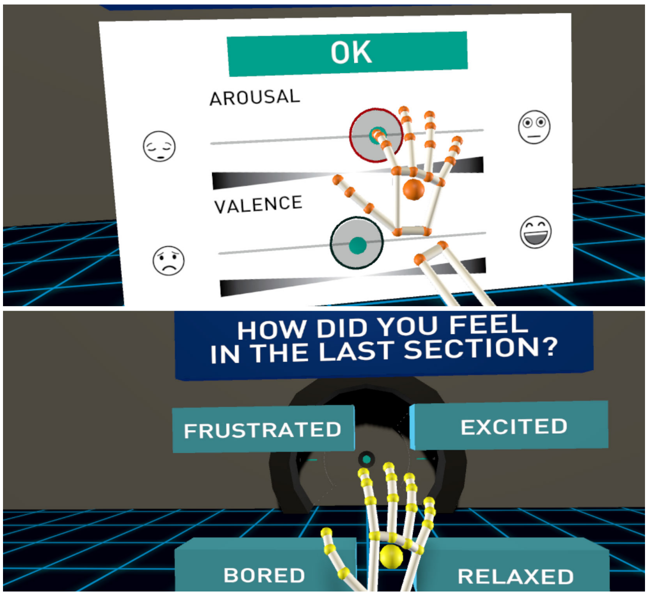

Figure 5.6: Two ways of self-reporting affective states tested in the pilot studies: 4 emotions (bottom) and the Affective Slider (top)

in Weka. Due to the computational complexity of some features (i.e.: peak detection), all features were computed in real-time in MATLAB when requested by the game. Given that features for machine learning had to be calculated in real-time, it was difficult to replicate the same results obtained in Study 2 with the application HRVLogger [5] for the HRV feature LF/HF ratio. Hence, this feature was removed from the HR feature set, leaving the HMD+HR model for arousal classification with 35 features.

5.3 Pilot studies

Two ways of self-reporting the affective states were explored in three pilot studies (see Fig. 5.6). In the first method, participants had to choose among 4 proposed emotions (bored, relaxed, frustrated or excited) as these were the affective states detected by the affect recognition system. Participants reported that they struggled to choose between the proposed emotions as these categories were too restrictive. For example, some participants expressed they would never report boredom as they are actively playing a game they are interested in. Thus, a second method of self-reporting affective states was implemented. This method is a VR replica of the Affective Slider [21], the affective state self-report method used in Study 2. This VR replica had the same controllers and graphics as in the original version designed by Betella et al [21]. Both methods were tested in the second pilot study. Participants reported to have more flexibility using the Affective Slider to report their affective states, so it was selected for the final study. Moreover, using this method it would be possible to assess the machine learning accuracy by comparing its output with the reported levels of arousal and valence in each section.

The real-time adaptation was tested with three pilot studies with 15 participants in total. The SVM parameters were fine tuned to improve the classification accuracies. The results showed a poor classification accuracy for both arousal (57%) and valence (47%). The number of features in the valence model was increased to the top 10 ranked features selected by correlation. This improved the valence classification accuracy to 57%, which is still close to the chance level of 50%. New parameters were set for arousal and valence SVM algorithms, achieving accuracies of 69% and 65% respectively. Since the affect recognition accuracy was not sufficient to successfully adapt the game by itself, a second decision layer was added based on players' performance. This approach has been suggested by others to improve the adaptation performance [25]. As most participants expressed that their affective states were strongly dependent on their performance, the score achieved in each section was correlated with the reported arousal and valence, finding a significant positive correlation with the later (rho=0.34, $p<.01$).

5.4 Decision rules for adaptation

In order to create a performance-based decision layer, the average score obtained in the sections where participants reported to be frustrated (i.e.: high arousal and negative valence) during the pilot studies was calculated. A mean score of 52 points (SD: 38.32) was used as a threshold to decide whether the player was achieving a high or a low performance (see Fig. 5.7). Thus, a two-layer decision system was designed using the affect recognition's output and the score (performance) achieved in each game section. Similarly to Afergan et al [3], two buffers stored the last 4 arousal and valence classification, keeping a time series of previous predictions. Since each *Memory Break* play consists of 5 sections, the researcher empirically decided to create buffers of width 4 to perform adaptation decisions when the engine is confident about the detected affective state (buffer mean $>.05$). This would act as a filter to smooth misclassifications and prevent erroneous adaptation decisions based only on the last classification prediction. A similar buffer stored whether the last four scores achieved were above (1) or below (0) the frustration threshold calculated. The average of these buffers was calculated for adaptation decision making.

An adaptation decision logic was designed where the difficulty level could be

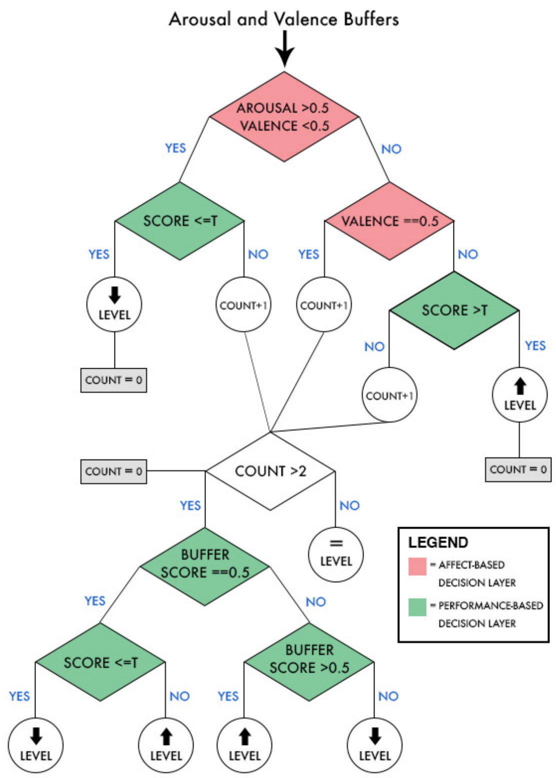

Figure 5.7: Adaptation decision logic of *Memory Break*. Note the two-layer decision system based on affect (arousal and valence) and performance (score).

increased, decreased or sustained (see Fig. 5.7). If the difficulty was sustained for more than 2 sections, a decision was taken based on the score buffer. At the end of each game play, the last difficulty level played and all buffers were saved for the next game play or session.

5.5 Summary

This chapter described the development and testing of an adaptation engine built from two decision layers based on affect and performance metrics. An affect recognition system was created to detect the player's arousal and valence levels using physiological and motion sensors. The detected affective state feeds the affect-based decision layer to make an adaptation decision. Due to the poor performance of the affect recognition system in a series of pilot studies, a performance-based decision layer was added as a second layer in the adaptation engine. This decision layer analyses the player's score in each game section and makes a final adaptation decision to increase, decrease or sustain the game's difficulty level.

The next chapter presents a longitudinal study where the affect recognition system and the adaptation engine were tested in real-time. Two versions of the game *Memory Break* were built, one with and one without the adaptation engine. The aim of this study is to explore the effects of difficulty adaptation in the player's WM performance when playing a VR game for WM training.

Chapter 6

Study 3: Real-Time Difficulty Adaptation for WM Training in VR Gaming

This chapter describes the last study of this research. This study explores the effects of adaptation in a VR game to train WM performance. Two versions of the game *Memory Break* were created, with and without the adaptation engine. The aim of the adaptation is to keep players in an optimal affective state and engagement level to improve their WM performance. The adaptation is determined by a two-layer decision system that changes the game's difficulty in real-time depending on the inferred affective state and the score achieved. Using a number of features extracted from physiological and motion sensors, two machine learning classification algorithms were trained and implemented in *Memory Break* to detect low and high levels of arousal, as well as positive and negative valence. In addition, this study investigates the impact of adaptation in the player experience and the effects of the players' self-reported affective states in their WM performance, following the findings of Study 2.

This chapter is structured as follows: section 6.1 describes the tasks and procedure of this study. A new version of *Memory Break* with the adaptation engine implemented and the difficulty adaptation are explained in section 6.2. Next, section 6.3 informs about the participants of this study and the questionnaires and sensors used, as well as the data pre-processing. Section 6.4 reports the performance and accuracy of the adaptation engine and the results and analyses made. Finally, section 6.5 discusses the findings of this third and final study.

6.1 Research design

The research design of this study is very similar to the previous study (Chapter 4). The same questionnaires and tests were used in this study to assess players' WM capacity, affective states and their player experience. In contrast to the previous study, participants played *Memory Break* only in the VR setting, ignoring the Desktop.

6.1.1 Tasks and setup

A longitudinal study was designed to evaluate the effects of VR game playing and difficulty adaptation on WM performance.The study consisted of 3 sessions where participants played the game 3 times in each session for approximately 5 minutes each play (see Fig. 6.1). Participants had to complete the 3 sessions within one week, leaving 2-3 days between sessions and booking all sessions at the same time of the day to minimise differences that might result from being tested at different times of the day. Participants were anonymised using IDs and compensated with £30 cash for their time and effort at the end of the study.

Similarly to the previous study, participants' WM capacity was assessed using the Automated Operation Span Task (AOST) test [56], described in section 4.2.1. Based on the score achieved in this test, participants were divided in two groups with high and low WM capacity. The game interaction was the same as in the previous study, using a hand tracking device (Leap Motion) to interact with the game. Moreover, participants had to wear the HR sensor (Polar H7) and an EMG armband (MYO) to measure their physiological signals for affect recognition.

6.1.2 Procedure

In their first session, participants had to complete a questionnaire about their background, gaming habits and motivations. Participants then completed the AOST test [56] to measure their WM capacity baseline. Before playing *Memory Break*, participants played a VR demo called *Blocks* to reduce the novelty effect the VR experience could cause. Participants interacted with the game using a hand tracking device (Leap Motion), and wore a HR sensor (Polar H7) and an EMG armband (MYO). At the start of each session, participants were asked to relax for 1 minute to record their HR baseline for later normalisation. Participants then practiced *Memory Break* before playing the game three times. Participants reported their arousal and valence levels throughout the game using the Affective Slider [21]. After each play, participants reported their level of engagement, motivation and motion sickness and completed the In-Game module of the Game Experience Questionnaire (GEQ) [66]. When all 3 plays were completed, participants reported their overall motivation, engagement and presence levels during the current session and completed the Post-Game module of GEQ [66], followed by a semi-structured interview about their experience and performance. Finally, in order to compare the effects of the adaptive and non-adaptive versions of *Memory Break*, participants completed the AOST again at the end of the third session.

6.2 Memory Break: Version 2

The logic, structure and interaction of the second version of *Memory Break* is the same as the first version, reported in section 4.3. However, due to the introduction of the adaptation engine, some changes had to be done to gradually adjust the difficulty level in real-time.

The goal of *Memory Break* is to obtain the highest possible score. The game consists of throwing balls at different stationary or moving obstacles to successfully pass through without crashing into them. If the player crashes, ten points are deducted from the score; one point is also subtracted every time a

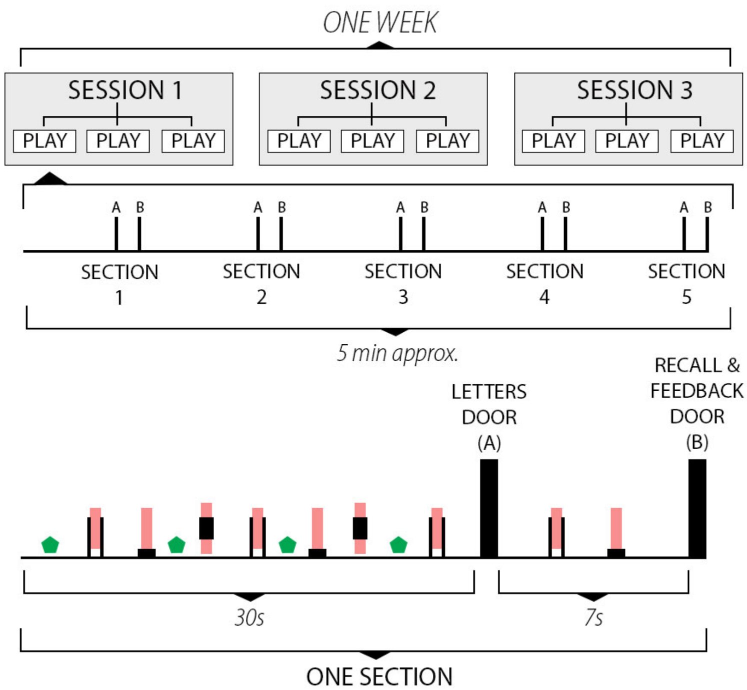

Figure 6.1: Structure of the present study and section division of a *Memory Break* play

ball is thrown. In order to get points, the participants have to throw and hit the gems found on their way, which added 10, 20, 30 or 40 points depending on the type of gem collected. As in the first version, each game play is divided into five sections consisting of 30s of game playing and a WM trial (see Fig. 6.1). The WM trial is structured as follows: the game stops at a door where a random sequence of letters appears that has to be remembered. These letters appear one at a time for 800ms, with 500ms gaps between them. After the sequence is displayed, the doors open and the game continues for another 7s. The game then stops again at another door where the player has to recall and input the letters previously shown in the exact same order. Once completed, feedback is provided indicating how many letters have been recalled correctly, the doors open and the game continues to the next section. In order to encourage players to remember the letters, five points are given for every letter recalled correctly. This decision was taken to connect the WM trials and the game, as some participants of the previous study saw them as two separate tasks. Before displaying the letters in each section, participants reported their arousal and valence levels using the Affective Slider [21]. The duration of each play is approximately five minutes. A score board with the ten best scores was displayed on one side before starting to play to motivate players to achieve a good performance.

6.2.1 Difficulty adaptation

Since the adaptation engine dynamically adapts the game's difficulty level, it was necessary to introduce some changes on how the game is generated. In contrast to the game's first version, this new version procedurally generates the sections in real-time depending on the difficulty level decided by the adaptation engine. *Memory Break* consists of 10 levels that gradually increase the difficulty. The difficulty is manipulated increasing the speed (+2%), the number (+3) and the type of obstacles presented. The types of gems also changed, placing those with more points in the harder levels to keep a balance between points collected and difficulty. As in the first version, the number of obstacles between doors also increased by two units every two difficulty levels. These changes maintain a similar difficulty in levels 1, 5 and 10 than in the easy, medium and hard levels of the previous study. Since the game has to gradually change the difficulty in a subtle way and without the player noticing, these changes were necessary to avoid big gaps between adjacent difficulty levels.

One of the most important changes to this new version of Memory Break is the WM trials. The length of the letter sequence presented in the WM trials changes depending on the player's WM performance. If all letters are recalled correctly, the sequence length is increased by one letter, otherwise it would stay the same. If the player fails to successfully recall all letters two consecutive times, the sequence length is reduced by one letter [170]. Whilst the sequence length always start with 3 letters in every session, the sequence length is saved between plays, reducing one letter when starting the next play. This method of changing the letter sequence length allows to adapt the difficulty of WM trials to the players' performance, pushing their limits and constantly challenging their WM.

The only difference between the adaptive and non-adaptive versions of *Memory Break* is the difficulty level adaptation. While the adaptive version changes the game's difficulty depending on the decision of the adaptation engine, the non-adaptive version constantly increases the difficulty level. In both versions, the difficulty is reduced by 2 levels if the player finishes the previous play of the current session in level 5 or above. Otherwise, the difficulty is reduced by 1 level. All sessions of the non-adaptive version always start in the first difficulty level. Conversely, the adaptive version loads the last difficulty level played in the prior sessions, applying the aforementioned rules. The administration of the WM trials is the same in both game versions.

6.3 Data collection

This study used the same data collection methods employed in Study 2. Quantitative and qualitative data was gathered for a better understanding of the impact of adaptation and the players' experience throughout the three sessions. This section describes the questionnaires and sensors used as well as the data pre-processing.

6.3.1 Participants

Fourteen participants, seven male and seven female, with a mean age of 26.78 (SD: 2.64) took part in the study. Seven participants were randomly assigned to the adaptive version of *Memory Break* and seven to the non-adaptive, keeping age and gender balanced. None of the participants had been diagnosed with any learning difficulty such as Dyslexia and do not experience motion sickness easily (i.e.: when reading while traveling as a passenger in a car). 71% of the participants had experienced VR before the study. 29% of the participants reported to have played video games between 0 and 2 hours the week before the study, while 43% did not play any. The motivation to play video games of ten participants was just for fun and only three reported to play to get the best score or be the best player. All participants in the adaptive version liked the game, except for one in session 1. In a similar manner, another participant in the non-adaptive version did not like the game in sessions 2 and 3, reporting boredom. Some participants found the interaction difficult to control, more specifically, eight in the first session, four in the second and five in the third.

6.3.2 Questionnaires and sensors

The questionnaires used in this study were the same as in Study 2: the pre-experiment, the post-game (or post-condition) and the post-session questionnaires. First, the pre-experiment questionnaire (see Appendix B.2) was completed at the beginning of the study. This questionnaire gathered participants' demographic data as well as gaming habits and experience. After each game play, participants self-reported their levels of immersion, engagement and motion sickness in the post-game questionnaire (see Appendix D.2). This questionnaire also included the In-Game module of the Gaming Experience Questionnaire (GEQ) [66], which measures 5 components of the player experience: Competence, Flow, Sensory and Imaginative Immersion, Tension, Challenge and Positive and Negative Affect. Participants self-reported their arousal and valence levels using the Affective Slider [21] in each section of the game (see Fig. 6.1), instead of at the end of each game play as in Study 2. Finally, at the end of each session, participants completed the post-session questionnaire (see Appendix D.3) about their overall experience in the current session. Participants' WM capacity was measured using the AOST at the beginning and the end of the study to assess improvements on their capacity.

The sensors employed were also the same as in the previous study. Participants wore a chest strap HR sensor (Polar H7) and an electromyograph (MYO) to measure their physiological signals for affect recognition. Whilst the HR data was gathered and pre-processed off-line by the app HRVLogger [5] in Study 2, the introduction of the adaptation engine in this study required a real-time collection of the physiological signals. Since this was not possible with HRVLogger, a small application was developed to record the HR sensor data and stream it to a computer via Bluetooth in real-time. This data was processed by MATLAB when required by *Memory Break* to make adaptation decisions. Due to the rapid technological advancements in VR, the HMD used in the previous study (Oculus Rift DK2) was no longer supported by Unity, the software used to create *Memory Break*, so it was replaced by a HTC Vive headset. The interaction with the game was the same as in Study 2, using the hand-tracking device Leap Motion (LM) and the same gestures for interaction.

6.3.3 Data pre-processing

The data gathered in this study did not require a great deal of pre-processing. Since this study was interested in the effects of adaptation in players' experience, WM and affective states, no physiological or motion analysis was carried out.

Only the WM measures required some pre-processing. Prior to the analysis of participants' WM capacity, measured with the AOST, the scores were normalised by dividing the number of letters correctly recalled by the number of letters presented (25). Thus, the percentage of letters correctly recalled was obtained. However, the analysis of participant's WM performance within the game used the number of letters correctly recalled instead of the percentage, since the number of letters displayed changed depending on the player's WM performance (see section 6.2.1).

6.4 Analysis and results

This section analyses the results of this last study. The analysis investigates the influence of the dynamic difficulty adaptation on players' WM performance, affective states and their player experience. Following the findings of the previous study (Chapter 4), this section also analyses the effects of valence and arousal on players' WM performance. Due to the small number of participants, the analysis focuses not only on statistically significant results but also on particular cases and individuals, analysing quantitative and qualitative data.

6.4.1 Statistics

The statistical methods used in the analysis of this study were similar to the methods used in Study 2 (Chapter 4). First, Levene's tests were carried out with the independent and dependent variables of interest to check for homogeneity of variance (homoscedasticity). Since the p-value of these tests were above the significance level (0.05), the homogeneity of variance was assumed. Variables were also plotted in histograms and Q-Q normal plots to check its distribution. Within and between repeated measures analysis of variance (ANOVAs) were performed to investigate significant differences between sessions and game versions. Since some data points from certain participants and sessions were missing due to technical problems, Linear Mixed Effects (LME) models were employed to perform analysis of variance (ANOVAs) as they allow for missing data points [168]. LMEs have been used in many studies for the analysis of self-reported and continuous measures in gaming research [128]. Further post-hoc t-tests were conducted where relevant. LME models were also used for regression analysis using random and fixed effects. Likelihood ratio tests were conducted to evaluate the goodness of fit of the models built. Finally, Spearman's correlations (rho) were carried out to estimate the relationship between two variables of interest.

Due to multiple comparisons in certain correlations and t-tests, Bonferroni corrections [1] were applied where one variable was tested against more than one condition. New significance levels after corrections are reported.

6.4.2 Classification accuracy and real-time adaptation

To assess the real-time performance of our classification models, the same thresholds used for training to divide the self-reported arousal and valence into two classes were used for testing. However, participants in this study reported significantly lower mean levels of arousal ($t=6.22$, $p<.001$) and valence ($t=3.99$, $p<.001$) compared to our previous study. Whilst the mean self-reported arousal was 0.87, in the current study it was 0.75. Similarly, the mean self-reported valence was 0.85 in the previous study and 0.76 in the current study. This means that participants in this study used a bigger range of values when reporting their affective states compared to participants in Study 2. One session of two participants were removed from this analysis due to technical problems with the sensors and the Affective Slider. Using these thresholds, the distribution of arousal classes was 220 low and 65 high arousal data points. Valence had a similar distribution, with 214 negative and 71 positive data points.

Two SVM algorithms were used for arousal and valence classification. Whilst the arousal classification model was trained with 35 features extracted from the HMD and HR sensors, the valence classification model was trained with 10 features from the HMD and EMG sensors. Arousal classification achieved an accuracy of 41% and valence 42%, both below chance level (50%). Table 6.1 illustrates the confusion matrices of arousal and valence classification. These

[1]The Bonferroni correction is used to avoid Type I errors when making multiple comparisons. It is calculated by dividing the significance level (in this case 0.05) by the number of tests that are being performed.

Table 6.1: Confusion matrices of arousal and valence classification in real-time

		Arousal Reported	
Arousal Prediction		**Low**	**High**
	Low	51	7
	High	169	58

		Valence Reported	
Valence Prediction		**Negative**	**Positive**
	Negative	99	49
	Positive	115	22

Table 6.2: Accuracies of arousal and valence classification in each session for all participants playing the adaptive version.

Participant	**Affect. Dimension**	**Session** *1*	*2*	*3*	**Mean**
1	Arousal	-	.67	.67	.67
	Valence	-	.4	.4	.4
2	Arousal	.53	.67	.4	.53
	Valence	.4	.53	.4	.44
3	Arousal	.67	.07	.2	.31
	Valence	.53	.67	.13	.44
4	Arousal	.07	0	.13	.07
	Valence	.53	.2	.33	.35
5	Arousal	.93	.93	.8	.88
	Valence	.13	.2	.33	.22
6	Arousal	.4	.2	.13	.24
	Valence	.8	.53	.6	.64
7	Arousal	.27	.13	-	.2
	Valence	.53	.4	-	.47

results evidence a poor classification performance of the models, which led to erroneous real-time adaptation decisions of the affect-based decision layer, although the performance-based decision layer worked as expected. Due to the imbalance of classes (see Table 6.1), an additional analysis examined the precision and recall of both arousal and valence classification. While precision expresses the proportion of data points classified as relevant that were actually relevant, recall represents the percentage of relevant instances over the total amount of relevant instances. Arousal showed a good recall of 0.89 but a much lower precision of 0.25. The classification of valence got poorer results, with a recall of 0.31 and a precision of 0.16. F1 measures, which combines precision and recall, were also computed, being 0.40 for arousal and 0.21 for valence. These results demonstrate a poor real-time performance of the classification models. Since participants were different in the current and the previous study, significantly lower levels of self-reported arousal and valence were found in this study compared to our previous study (training stage), which could explain the imbalanced distribution of classes in the testing dataset.

Table 6.2 presents the arousal and valence classification accuracy of each participant in the adaptive version in every session. The best mean prediction accuracy of arousal was 88% for participant 5. This participant reported high levels of arousal ($>.92$) during the whole study that were successfully detected. The worst performance of arousal classification achieved a mean accuracy of 7% for participant 4, who's mean self-reported arousal was 0.68 (SD: .1). Using the thresholds applied during training, her self-reported arousal was always labelled as low ($<.87$) but wrongly classified as high arousal. Similar results were found for participant 3 in sessions 2 and 3. On the other hand, valence classification showed similar results. Participant 6 got the best mean valence accuracy with 64%. The overall mean valence reported by this participant was 0.63 (SD: .19), which was labeled as negative valence and successfully classified by the model. Similarly, participant 5 always self-reported high levels of positive valence (Mean: 0.97; SD: .07), although it was mostly predicted as negative valence, achieving the lowest mean classification accuracy (22%). These results, together with the confusion matrix table (see Table 6.1), indicate that the arousal and valence models mainly predicted high levels of arousal and negative valence. Nevertheless, looking at Table 6.1, most of the self-reported low arousal was successfully recognised.

Although the arousal and valence classification did not work well overall, the adaptation worked better for some participants depending on individual factors such as motivation. For example, participants 3 and 5, whose motivation was to be the best player, played in a more aggressive manner moving their head and hands very abruptly. This was mostly classified by the machine learning algorithms as frustration (high arousal and negative valence), which made the adaptation engine to sustain or reduce the difficulty level, keeping participants in the easiest levels (1-4). Due to individual differences in preferences or motivations to play video games, participants can experience video games differently or have different playing styles [142]. One of the participants, who never experienced VR before, reported having a positive experience when interviewed, even though the game was sometimes too difficult for her. This may be explained by the excitement of trying VR for the first time, known as novelty effect [96]. These individual differences present important challenges in the design of generic adaptation methods and subject-independent machine learning models.

6.4.3 Player experience

The analysis in this section is focused on the variables measured by the In-Game and Post-Game modules of Game Experience Questionnaire (GEQ) and other variables reported at the end of each game play such as engagement, motivation or motion sickness. The self-reported engagement, motivation or flow did not show any significant differences between game versions or sessions, although motion sickness was significantly different between sessions ($F=3.59$, $p<.05$), indicating that participants' motion sickness decreased along sessions. Immersion showed significant differences between sessions ($F=4.07$, $p<.05$), being slightly more pronounced in the adaptive ($F=2.53$, $p=.09$) than in the non-adaptive ($F=1.88$, $p=.16$) game version. A further post-hoc analysis between sessions revealed significant ($p<.02$, Bonferroni correction applied) differences between sessions 1 and 2 in the adaptive version ($t=-2.41$, $p<.02$). The negative t-value demonstrates a significant decrease in immersion in session 2 in the adaptive version. The self-reported competence showed almost significant differences between sessions in the non-adaptive version ($F=2.75$, $p=.07$), and in the interaction between game version and sessions ($F=2.39$, $p=.09$). Significant differences were also found between sessions in tension ($F=7.97$, $p<.001$) and challenge ($F=4.71$, $p<.01$). A post-hoc analysis in the self-reported challenge only showed significant differences between sessions in the non-adaptive version ($F=6.19$, $p<.01$). No significant differences were found between game versions or sessions in the reported positive affect, measured by the GEQ. However, negative affect was almost significant between sessions only in the adaptive version ($F=2.84$, $p=.06$), concretely between sessions 1 and 3 ($t=-1.87$, $p=.07$).

The analysis of the Post-Game module, completed at the end of each session, showed more interesting results. Significant differences were found in the self-reported positive experience only in the adaptive version ($F=3.86$, $p<.05$). A further paired t-test analysis showed significant differences between sessions 2 and 3 ($t=3.20$, $p<.02$), indicating that participants had a significantly higher positive experience in session 3 compared to session 2. On the other hand, the reported negative experience was almost significant in the non-adaptive ($F=2.96$, $p=.09$) but not in the adaptive version ($F=0.90$, $p=.43$). A further analysis revealed a decrease in negative affect in session 2 ($t=-2.11$, $p=.08$), followed by an increase in session 3 ($t=2.66$, $p=0.04$).

These results evidence a decrease in immersion, tension and challenge along the sessions in both game versions. This can be explained by the repeated exposure of participants to the game, being less immersed and challenged throughout the 3 sessions [96]. However, participants in the adaptive version had a greater positive experience in session 3 than those in the non-adaptive version, who reported higher levels of negative experience in the last session. This could be explained by the game's adaptation, which may have had a positive effect in the player experience of those playing the adaptive version, although further research needs to be done to clarify these effects.

Almost all the variables measured by the In-Game module of the GEQ correlated significantly ($p<.008$, Bonferroni correction applied) with the mean self-reported arousal and valence of each game play (see Table 6.3). As expected, the only two variables that had a negative correlation were tension and negative affect.

Table 6.3: Correlations of self-reported arousal and valence with the In-Game module of the GEQ.

Param	**Dependent Variable**	**Rho**	**Sig.**
Arousal	Immersion	.33	<.001
	Flow	.19	.04
	Competence	.68	<.001
	Challenge	.23	.01
	Tension	-.29	<.001
	Negative Affect	-.22	.02
	Positive Affect	.53	<.001
Valence	Immersion	.33	<.001
	Flow	.24	<.008
	Competence	.71	<.001
	Challenge	.23	.01
	Tension	-.26	<.008
	Negative Affect	-.17	.07
	Positive Affect	.51	<.001

6.4.4 WM performance

Within-subject repeated measures t-tests were carried out to analyse the scores achieved in the AOST test before and after the study. One session of two participants in the adaptive version were removed from this analysis as the game had to be rebooted. Participants in the non-adaptive version showed almost significant differences (t=2.15, p=.07), but not those in the adaptive version (t=0.46, p=.66). Regarding participant's WM performance within *Memory Break*, no significant differences were found between sessions in the adaptive (F=2.61, p=.12) or non-adaptive (F=0.53, p=.60) versions. However, a further post-hoc analysis showed significant (p<.02, Bonferroni corrections applied) differences between sessions 1 and 3 in the adaptive version (t=2.26, p<.02) but not in the non-adaptive (t=0.95, p=.36). Additionally, significant differences were found in the maximum number of letters correctly recalled per session in the adaptive (F=4.78, p<.05) but not in the non-adaptive version (F=1.11, p=.36). These results indicate that participants in the adaptive version had a greater improvement in their WM performance than those in the non-adaptive version.

Spearman's correlations were also conducted between the number of letters correctly recalled in each game play and the self-reported player experience variables measured by the In-Game module of the GEQ. Participants in the adaptive version showed strong significant correlations with immersion (rho=0.65 p<.001), competence (rho=0.57 p<.001) and positive affect (rho=0.42, p<.001), whilst those in the non-adaptive version showed a weaker but significant correlation with competence (rho=0.33 p<.01) and positive affect (rho=0.27 p<.02), and no significant correlation with immersion (rho=0.20 p=.12). Similar correlations were found with the maximum letters correctly recalled in each game play. Overall, these results suggest that participants had a better WM performance when they felt more immersed and had higher levels of positive affect, feeling more successful in the game. These effects were more pronounced on participants in the adaptive version. These findings are related to the flow theory [45], which proposed that cognitive performance can improve when subjects experience positive affective states.

Following the same approach as in Study 2 (Chapter 4), participants were divided in two groups (low and high WM capacity), using the overall median score (0.79) of participants' WM capacity as measured with the pre-study AOST. Seven participants were assigned to each group. Significant differences were found between the pre- and post-study AOST scores in the low WM group (t=2.69 p<.05) but not in the high WM group (t=-0.40 p=.70), who actually had a slightly worst performance in the post-study AOST.

Participants in the low and high WM groups were subdivided depending on game version played. Significant differences were found in the maximum number of letters correctly recalled in the low WM group in the adaptive version (F=6.71 p<.05) but not in the non-adaptive version (F=0.18 p=.84). Participants with high WM in the adaptive version showed a greater but not significant difference between sessions (F=3.48; p=.17) than those in the non-adaptive version (F=0.39; p=.69).

6.4.5 Affective states and WM performance

This section investigates the effects of the self-reported valence and arousal in WM performance, following the findings of the previous study: high levels of arousal and positive valence can have positive effects on the players' WM performance. One session of two participants were removed from this analysis as they had problems with the Affective Slider or deliberately manipulated it, setting the sliders to the maximum "*to see what happens*", as reported to researchers. The reported arousal and valence of players in the adaptive version correlated very significantly (p<.001) with the number of letters correctly recalled, both with a correlation coefficient of 0.28. Similar results were found in the non-adaptive version for arousal (rho=0.23, p<.001), though valence had weaker correlation (rho=0.14, p<.05).

Figures 6.2 and 6.3 show the self-reported arousal and valence in every session and game play of participants with low and high WM capacity, together with their WM performance (number of letters correctly recalled). As expected, participants with high WM capacity correctly recalled more letters than those

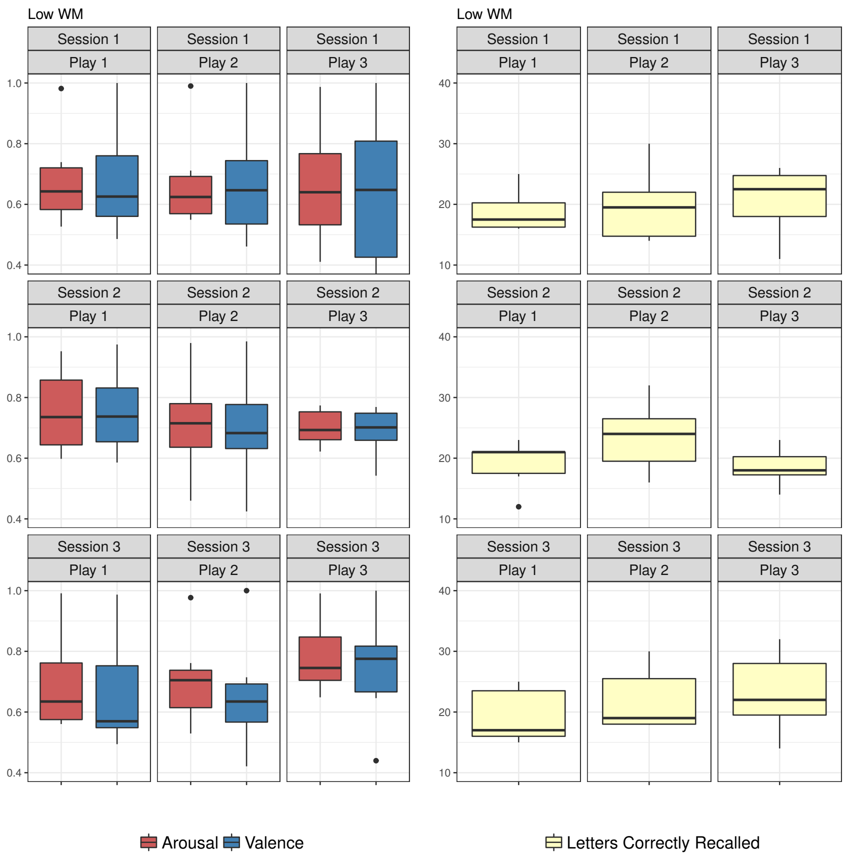

Figure 6.2: WM performance (right) and self-reported arousal and valence (left) of participants with low WM capacity. Participants in this group got their best WM performance in play 3 of session 3, which corresponds with their highest levels of arousal and valence, especially the later.

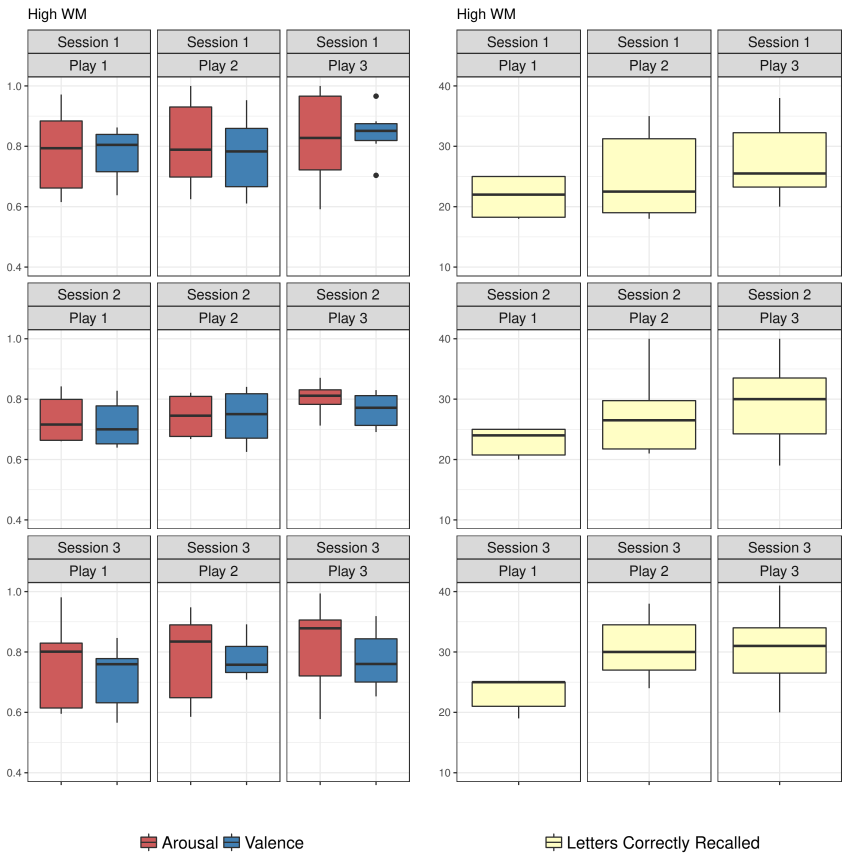

Figure 6.3: WM performance (right) and self-reported arousal and valence (left) of participants with high WM capacity. Participants in this group got their best WM performances in plays 2 and 3 of session 3, which corresponds with their highest levels of arousal but not of valence.

with low capacity. According to the results obtained in Study 2, high levels of arousal and valence may have positive effects on players' WM performance. However, as it can be seen in Figure 6.3, this is not true for subjects in the high WM group. Participants with high WM capacity achieved their best WM performances in plays 2 and 3 of the session 3, reporting their highest levels of arousal but not of valence. Similarly, participants with low WM capacity achieved their best WM performance in play 3 of session 3, which corresponds with their highest levels of self-reported arousal and valence of the whole study.

In order to further explore the relationship between affective states and WM performance, the mean self-reported arousal and valence per game play were correlated with the percentage of letters correctly recalled in each game play. The percentage instead of the number of letters correctly recalled was used as it describes how many of the given letters were successfully recalled instead of how far they got up to in terms of letters recalled. Participants with low WM capacity showed significant moderate correlations for both arousal (rho=0.40, $p<.01$) and valence (rho=0.29, $p<.02$), whilst participants with high WM capacity did not show significant results (arousal rho=0.15, $p=.24$; valence rho=0.16, $p=.23$). These results suggest that valence and arousal might have stronger effects on the WM performance on those with low WM capacity compared to subjects with high WM capacity.

Finally, Linear Mixed Effects (LME) regression models were used to predict participant's WM performance, this time measured as the percentage of letters correctly recalled. A model was created using the self-reported arousal and valence as fixed effects and subjects as random effects. Arousal and valence were the only predictors as this research is interested in their effects in participants' WM performance. As seen in Table 6.4 only arousal was a significant predictor. A likelihood ratio test was performed, showing the goodness of fit of this model (Chi-sq=10.39, $p<.01$).

Table 6.4: LME model of WM performance

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.63	0.06	94.66	10.11	0.00
Valence	0.03	0.09	559.18	0.27	0.78
Arousal	0.24	0.11	343.00	2.31	0.02

6.4.6 Qualitative analysis

After each session, a semi-structured interview was conducted to understand how participants felt in terms of player experience and performance. Some participants reported not caring much about doing well in the WM trials within the game and just tried to remember up to a certain number of letters, even if more letters were presented. For example, some participants assumed they could not recall more than 6 letters, so even if more letters were presented, they only tried to remember the first 6 letters as they could not remember more. This attitude might explain why some participants reported paying less attention to the WM trials within the game when they were focused on achieving a good performance in the game (i.e.: trying to get a high number of points). This could be explained by the fact that some participants did not see the WM trials as part of *Memory Break* but as a separate task within the game. As reported by some participants, if they suffered memory or learning difficulties, they would then be more interested in doing well in the WM trials as it would be beneficial for them.

The Score Board displaying the top 10 scores proved important as a method of keeping players interested in the game, especially for those whose motivation was to get the best score. Moreover, some participants whose motivation was to play '*just for fun*', reported that the Score Board encouraged them to get a better score, making them feel more challenged. Most participants reported that failing to achieve their goals or getting a low score led them to frustration or stress and thus, to have a bad WM performance. Conversely, when participants experienced high arousal levels accompanied by high levels of positive valence, associated with feelings of being excited and successful, they reported to have a better WM performance. As one participant reported, "*being super excited does not help with memory, you have to be excited but not too much*".

Some of the errors participants made in the WM trials within *Memory Break* was due to confusions with letters that are graphically or phonetically similar. Furthermore, participants sometimes missed the first letter of the sequence displayed and, consequently, they got all the sequence wrong as they did not input all letters in the exact same order as displayed.

6.5 Discussion

6.5.1 Classification and adaptation performance

The poor accuracies of the classification models could be explained by the significantly lower mean levels of self-reported arousal and valence in this study. Due to these differences between the machine learning training and testing, the distribution of classes in the real-time testing dataset was imbalanced, as the thresholds used during training were employed to divide arousal and valence into 2 classes each. One of the reasons that might explain these significant differences in the self-reported arousal and valence between studies is the repeated exposure to our game. Previous studies have demonstrated that physiological response intensity decreases after the first exposure when a subject experiences something novel [96]. Even though physiological signals were normalised in every session, this effect, called novelty effect, could have decreased the self-reported arousal and valence levels of some participants, leading to substantial differences in the class distribution.

Since the difficulty adaptation engine comprised two decision layers (see Fig. 5.7) the performance-based decision layer corrected some incorrect decisions taken by the affect-based layer due to misclassifications of arousal and valence. For example, when the affect classification algorithms erroneously predicted frustration (high arousal and negative valence), the performance decision layer corrected incorrect adaptation decisions analysing the score achieved. However, the performance-based decision layer could be substantially improved as there was important differences between subjects about what merited high performance. Another relevant aspect of the adaptation engine was the creation of buffers for arousal and valence, which avoided reactionary adaptation decisions that could happen due to arousal or valence misclassifications. These buffers worked in a similar way to a mean filter, smoothing the impact of erroneous classifications on the adaptation decisions averaging the last 4 predictions. Since misclassifications are unavoidable, it is important to know when and how adaptation decisions should take place to provide an optimal experience [3].

These results suggest that thresholds may not be appropriate for machine learning labelling and classification due to individual differences between subjects. For example, using a static threshold to divide participants' self-reported arousal in 2 levels (high vs low) can be problematic since not all subjects would report similar levels of arousal or use the same range of values with the Affective Slider. Thus, further research needs to be done to find methods to detect and overcome individual differences, avoiding the use of static thresholds for all subjects. One suitable solution would be to normalise participants' self-reported affective states. Another important question in the design of machine learning models is the use of subject-dependent or subject-independent classification models. As shown in previous research [119], subject-dependent models (i.e.: training and testing a model with the same subject) can be more accurate than subject-independent models. Thus, the development of subject-dependent classification models for adaptation can bring more accurate adaptation and personalised cognitive training, although they are more expensive to produce as they need to be specifically created for each user. Conversely, subject-independent classification models, as it has been done in this study, are more challenging to create as they should work for all subjects. As shown in Table 6.2, the performance of the models varies greatly between participants due to differences in their motivation and ways of playing, which changed their physiological and

Table 6.5: Number of data points for each difficulty level of the Low and High WM capacity groups

	Difficulty Levels								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
High WM	17	33	58	46	46	31	30	20	22
Low WM	15	50	76	45	42	29	26	14	12

motion signals.

The physiological and motion sensors selected for arousal and valence classification in this study can inform future research about effective ways to detect player's affective states in VR gaming. Whilst the HR sensor and HMD have been previously used for arousal detection [128][17], the sensors used for valence recognition show more interesting insights. The combination of the HMD and EMG armband suggest that valence could be inferred analysing the player's head motion and the force applied to the interaction controller when playing a VR video game. Nonetheless, it is important to reflect on why the hand's motion was not informative about the participants' affective states. One of the reasons could be the attachment of Leap Motion in front of the HMD. For instance, if participants held their hands still in front of the HMD, any head movement would be detected by Leap Motion as a hand displacement as it is moving together with the headset.

6.5.2 Adaptation and WM

Only participants in the adaptive version significantly improved their WM performance in the game's WM trials in the third session. However, these participants did not improve their performance in the AOST test. This could be explained by two factors: the complexity of the AOST test and the ordering of events (game play then AOST) in the final session. Since the AOST used math problems to be solved during the letters display, participants found it much harder than the WM trials within the game. Furthermore, participants completed the pre-study AOST at the beginning of session 1 and the post-study at the end of session 3, which produced different effects in certain participants. While some found the post-study more difficult than the pre-study AOST, due to tiredness of playing the game, others reported to be more prepared and active as they trained their WM playing *Memory Break* before the test.

Participants with low WM capacity significantly improved their score in the AOST test. Significant differences were also found in the maximum number of letters correctly recalled in subjects with low WM capacity in the adaptive version. These results suggest that adaptation can be more beneficial for those with low WM capacity. Participant's WM performance in terms of percentage of letters correctly recalled was also analysed along the difficulty levels. The results, as shown in Figure 6.4, suggest that while participants' WM performance decays as the difficulty increases, there seems to be a performance improvement in levels 7 and 8, this being more pronounced for participants with low WM capacity particularly in level 8. This finding can be linked to the flow theory, which suggests that optimal levels of challenge can lead to improvements in cognitive tasks when subjects feel successful. Thus, challenge and arousal levels elicited in difficulty level 8 may have helped participants with low WM capacity to have a better WM performance. The number of data points for each group is shown in Table 6.5. Results observed in difficulty level 10 were disregarded as the low WM group only had one data point and the high WM group had seven. More research would be needed to further investigate the effects of difficulty levels in WM performance.

6.5.3 Affective states, player experience and WM

The higher positive experience reported by participants in the adaptive version in session 3 could explain their significantly better WM performance in this session, although practice effects should be considered as participants were already familiar with the game from the last session. However, the significant correlations of self-reported immersion, competence and positive affect with their WM performance indicate that high levels on these variables may contribute towards a better WM performance. Moreover, some participants reported during the interviews that having a positive player experience (i.e.: getting a good score and feeling successful) led them to achieve a better WM performance. This is related to the optimal experience or flow theory [45]. While competence in-

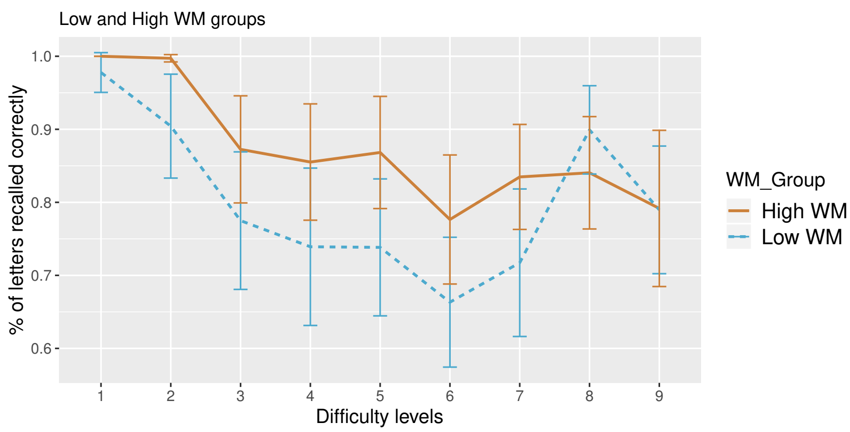

Figure 6.4: WM performance of participants with low and high WM capacity along difficulty levels of *Memory Break*

creases, participants feel more successful and immersed in the game, triggering positive affective states that can have positive effects in the players' cognition.

The significant correlation of self-reported arousal and valence with WM performance suggest that both can have relevant effects in participants' cognition. A further investigation revealed that participants in the low WM group seem to benefit more from high levels of arousal and valence than those with high WM capacity. This can be observed looking at play 3 of session 3 in Figures 6.2 and 6.3. The self-reported levels of arousal and valence of participants with low WM capacity may have helped them to achieve their best WM performance of the whole study in this session. However, the reported arousal and valence of the high WM group seems to not affect their WM performance as they got a good performance regardless of the affective state experienced. These findings are consistent with previous research reporting that subjects with high WM capacity have better self-regulation of emotional expressions [144].

LME models were used to predict participants' WM performance, using the self-reported arousal and valence as predictors. The results showed self-reported arousal as the only significant predictor of participants' WM performance, suggesting it has greater impact than valence. This was also shown in the correlations of arousal and valence with participants' WM performance. Although both arousal and valence correlated at the same level (rho=0.28) with participants' WM performance in the adaptive version, the correlation of valence of participants in the non-adaptive version was much weaker (rho=0.14). These results are in line with previous studies suggesting that arousal has positive effects on our cognition up to a certain point [88], whereas the effects of valence are still controversial and might be task dependent [19].

6.6 Summary

The study reported in this chapter draws together the work undertaken in this research to highlight the importance of adaptation in game-based cognitive training programs as well as the effects of affective states on cognitive performance. An adaptation engine, composed of affect and performance-based decision layers, was designed to dynamically adapt the difficulty of *Memory Break*. Participants playing the adaptive version of the game significantly improved their WM performance in the game's WM trials in session 3, though no significant differences were found between the pre and post-study AOST test. The reported positive experience and the significant positive correlations of immersion, competence and positive affect with WM performance indicate a positive effect of these variables in participants' WM performance. The results also suggested that arousal has stronger effects on participants' WM performance than valence. Following these findings, this research suggests that positive player experiences can improve the player's cognitive performance. Thereby, real-time adaptation in game-based cognitive training is crucial to achieve effective training and real transfer effects [8].

In addition, the findings indicate that arousal and valence have greater effects in participants with low WM capacity than those with high capacity. This suggests that affect-based adaptation could have positive results on subjects with low WM capacity as the affective states have stronger effects on their cognitive performance. Therefore, adaptation to keep players in an optimal affective state can be more beneficial for those with low WM capacity. These findings are important to improve not only cognitive training programs but other programs involving cognitive processes such as cognitive rehabilitation.

Chapter 7

Discussion and Conclusions

This last chapter summarises the principal findings and contributions of this research. The chapter is structured as follows: section 7.1 discusses the most important results of the three studies undertaken in this research, putting the main findings in context with previous relevant research. In section 7.2, the key findings and contributions of this PhD research are summarised. Section 7.3 documents the main limitations of the studies reported in this research. Section 7.4 proposes future research directions. Finally, this chapter concludes with reflections and lessons learned in carrying out research on affective gaming in VR.

7.1 Discussion and implications

There were three aims to this thesis: 1) to study the effects of VR gaming on player's WM performance and affective states, 2) to investigate the role of arousal, valence and immersion on player's WM performance, and 3) to explore how real-time adaptation and affective gaming in VR can be used for WM training. These aims were addressed in two studies (Chapters 4 and 6), in addition to an initial pilot study (Chapter 3) that examined the suitability of physiological and behavioural signals for affect detection in gaming. The second study was designed to investigate the effects of game playing in VR and Desktop on player's WM performance and affective states, as well as the role of these on WM performance. A final study examined how real-time difficulty adaptation could be used in VR gaming for WM training.

This section discusses the main findings of this research in the context of the literature reviewed. These findings are grouped under four topics: i) affect detection in gaming, ii) player experience, affective states and WM performance in VR, iii) effects of arousal and valence in WM, and iv) adaptation in games for WM training.

7.1.1 Affect recognition in gaming

During the three studies presented in this thesis, different physiological sensors have been used. An initial pilot study used ECG and GSR sensors to measure players' heart and electrodermal activity. ECG was a good and reliable sensor to evaluate the players' HR and HRV. Whilst HR can be used to measure arousal levels, as demonstrated in studies 1 and 2 as well as previous research [139, 127], HRV can serve as an indicator of the players' engagement levels. This finding is in line with previous studies finding HRV a good measure of engagement and cognitive load [155, 175].

One important aspect of using physiological sensors for affect detection in gaming is how and where the sensor is attached to the player. Christy and Kuncheva [41] claimed that standard sensors attached to the player's fingertips are not appropriate for active game playing as it can limit the player's movements or actions [81]. This was the case with GSR in Study 1. Placed on the hand holding the Wii controller, the GSR signal was extremely noisy due to constant hand movements. Similar results were reported by Morgan et al [105] when measuring physiological signals in collaborative music making. Sykes and Brown [154] also found GSR problematic for affect detection in gaming because skin conductance can increase when muscles tightens.

Therefore, not all physiological sensors are suitable for measuring affective states in gaming. Some participants of Study 1 found the ECG and GSR sensors very invasive since the electrodes had to be directly attached to their chest and fingers (see Fig. 3.2). These were replaced by wearable sensors in studies 2 and 3 as they offer a non-invasive and reliable way of measuring physiological signals without limiting the player's movements [41, 81, 65].

Behavioural signals have also been used for affect detection. Kleinsmith and Bianchi-Berthouze [76] claimed that arms and upper body show the most important behavioural cues. Specifically, head motion has been found one of the best features to detect affective states in different scenarios such as gaming to music performance [35, 54, 61, 17, 34]. The observational analysis of facial expressions, body postures, gestures and spatial behaviour in Study 1 demonstrated that it is possible to infer the player's affective states analysing these behavioural cues. Some researchers have suggested that body movements in game playing significantly depend on the player's motivation [143, 108]. Although the study of motivation in gaming is outside of the scope of this research, it is worth noting that motivations can change the way games are experienced [75]. For example, some people find certain negative emotions enjoyable (such as frustration), as these emotions form part of the gaming experience when the motivation is being the best player [58].

Although physiological and behavioural signals can be used to infer the player's affective states, various researchers have used the sensors in existing devices (i.e.: accelerometers in controllers) for multi-modal recognition of emotions [24, 65, 80]. Sykes and Brown [154] found a strong positive correlation between the player's arousal levels and the force applied to the gamepad. Study 2 explored whether this is true in hand gesture interactions, measuring the force applied to the hand with the EMG armband MYO. The results confirmed the suitability of hand's muscle activation to infer arousal levels in game playing, even if it is a gesture or motion-based interaction. A similar approach was undertaken to measure arousal levels analysing the player's head motion using the HMD's sensors. Yaw (Y rotation) and pitch (X rotation) were used by Becker-Asano et al [17] to infer subjects' arousal analysing their head motion through the HMD in a virtual emergency scenario. In line with their findings, the head's angular velocity was found to be a good indicator not only of arousal but also of valence. This indicates that the head's angular velocity can be used to detect the player's arousal and valence levels, although further research needs to be done to confirm its relationship with the latter.

In summary, physiological and behavioural signals can be used to measure players' affective states, although not all sensors are suitable as they can be too invasive. Physiological sensors that have to be attached to the player's body should not limit the player's movements or actions. Recent technological developments on wearable sensors allow a non-invasive, reliable and easy way of measuring physiological signals, suitable for gaming scenarios. Furthermore, motion sensors such as accelerometers or gyroscopes can be used to assess behavioural cues that inform about the player's affective states [75]. Thus, this research proposes the use of the motion sensors already present in many interaction devices such as HMDs or controllers for affect detection.

7.1.2 Player experience, affective states and WM performance in VR

One of the principal interests of this PhD research was to understand effects of VR gaming on players' experience, affective states and WM performance. Player's experience was measured with the Game Experience Questionnaire, which assesses different dimensions of the player experience such as immersion, flow, competence, tension or challenge [66]. This section discusses relevant findings of studies 2 and 3.

As described in Section 2.1.2, VR has been associated with high levels of immersion, often called presence [31]. Researchers have defined presence as the feeling of '*being there*' in the virtual world, detaching from the real world [131, 164]. Researchers have also referred to VR has an '*affective medium*' [131] or an 'experiential interface' [164] since it can directly influence the intensity of the experienced affective states. Riva et al [131] suggested a bidirectional interaction between presence and emotions. Whilst presence levels can intensify the affective states, the emotional responses in a VR environment can also increase the levels of presence. These findings are in line with the results of Study 2, where significant differences were found on the self-reported immersion, arousal and valence between VR and Desktop. However, the levels of immersion reported by participants of Study 3 decreased throughout the three sessions. This indicates that the intensification of immersion and affective states in VR can decrease as a result of multiple exposures. This finding is consistent with the research of Meehan et al [97], who found that participants' physiological responses decrease over multiple VR exposures. This is related to the novelty effect, which argues that subjects experience higher physiological responses when presented something novel [123]. This could explain why Pallavicini et al [117] did not find significant differences in participants' HR when playing *Smash Hit* (the game that inspired *Memory Break*) in VR and iPad. Although they argue these results could be due to the game chosen, it could also be due to previous exposure of participants to VR environments.

The effects of VR gaming in player's WM performance were also explored. The results of Study 2 revealed that participants had a significantly better WM performance in the easy and medium difficulty levels of VR but not in Desktop. A further analysis divided participants of Study 2 into two groups (low and high) according to their WM capacity. Only the low WM group showed a significantly better WM performance in these difficulty levels of VR. Similar results were found in the low WM group playing the adaptive version of *Memory Break*, whose WM performance and WM capacity improved. This demonstrates that adaptive cognitive training programs in VR can be more beneficial for subjects with low WM capacity, such as those diagnosed with Dyslexia or ADHD [78]. Since adaptive VR gaming for cognitive training have stronger positive effects on subjects with low WM capacity, as shown in this thesis, subjects ADHD would benefit more of this type of training as they normally suffer of low WM capacity [78]. However, further research needs to be done to confirm these effects on subjects suffering ADHD or Dyslexia. The better WM performance of participants in VR can be explained by the higher levels of immersion reported, which correlated significantly with participants' WM performance in studies 2 and 3. The significant correlation between the self-reported competence and WM performance in both studies suggests that competence also had a positive effect in participants' WM. Weinstein et al [166] suggested that feeling physically present in a virtual environment can augment the retention and transfer of content into the real world. Chittaro et al [39] found a higher knowledge retention of aviation safety instructions using a VR experience than a card game. Similar findings were reported by Olmos-Raya and colleagues [112] using a mobile-based HMD and a tablet for educational purposes. These findings indicate that high levels of immersion can have significant positive effects on subjects' cognitive skills.

In summary, participants experience higher levels of immersion in VR compared to other interactive modes. At the same time, VR intensifies the experienced affective states. These two factors can have significant effects on participants' WM performance, helping them to make a better use of their cognitive resources. These effects are stronger on participants with low WM performance. Thereby, VR can be an effective interaction mode for game-based cognitive training due to the positive effects of high immersion levels on cognitive skills.

7.1.3 Effects of arousal and valence on WM performance

As mentioned in the previous section, VR can significantly intensify the user's affective states. Due to the reciprocal influence between affect and cognition [164, 30], there was an interest in investigating the effects of arousal and valence on player's WM performance. This section discusses these effects in detail, reflecting on the results of studies 2 and 3.

To recap, Yerkes and Dodson [174] proposed that arousal has positive effects in cognitive skills up to a certain point after which it has negative effects if arousal continues to increase. Although the effects of valence are controversial [19], various studies have demonstrated that positive affective states can also benefit cognitive skills [140]. Olmos-Raya et al suggested that '*positive emotions stimulate curiosity, increase attention and arouse interest in the topic being learned*' [112].

Following these findings, Study 2 analysed participants' self-reported arousal and valence as well as their WM performance in each difficulty level. In order to have a clearer insight of the effects of arousal and valence on WM, this analysis was carried out on the low and high WM capacity groups independently. In accordance with Yerkes and Dodson, the highest levels of arousal corresponded with participants' worst WM performance. This finding is also in line with previous research that demonstrated arousal increase as a function of game difficulty [175, 158]. The effects of valence on WM performance were more evident in participants with low WM capacity. Their highest level of self-reported valence coincide with their best WM performance. These observations are consistent with the research of Curci et al [46] who suggest that the emotional valence of a stimuli and the individual's WM capacity influence the subject's WM performance. Schmeichel and colleagues [144] reported similar findings in a series of four studies, concluding that subjects with low WM capacity are less efficient in self-regulation of emotional experiences than those with high capacity. This indicates that subjects with low WM capacity can experience the detrimental effects of arousal on cognition earlier than those with high WM capacity.

Therefore, subjects with low WM capacity could benefit more of an optimal affective state when performing a cognitive task. Keeping them in a flow state where they successfully overcome the challenges presented, subjects with low WM can make a better use of their cognitive resources. Due to the effects of difficulty levels in subject's affective states, especially in arousal [175, 158], it is important to provide the right difficulty level to optimally challenge the player without causing frustration or boredom. Yeh et al [173] proposed a similar idea activating the subject's attention presenting adequate challenges and avoiding negative emotions that could detriment the cognitive performance. However, they did not find consistent effects of self-reported negative emotions on participants' WM, probably due to individual factors such as motivation or WM capacity.

Finally, arousal was found to be a better predictor of WM performance than valence. Linear Mixed Effects models were built in studies 2 and 3 to predict participants' WM performance using the self-reported arousal and valence as fixed effects. Whilst both arousal and valence were significant in Study 2, only arousal was significant in Study 3. This can explain why researchers have found important effects of arousal on WM performance, whilst the role of valence is still controversial [19, 174, 173].

7.1.4 Adaptation in games for WM training

Two main conclusions can be made so far: first, VR can be a useful medium for cognitive training due to the positive effects of immersion in subjects' cognitive skills, and second, arousal and valence can have significant effects in cognition. In light of these conclusions, this section discusses the findings of Study 3 regarding the adaptation engine, as well as the implications about why and how affective states can be used to adapt game-based cognitive training programs.

Video games have often been used to motivate subjects in low engagement educational or cognitive tasks [125, 160]. Since challenge is one of the most important aspects of video games [44], many researchers have suggested the implementation of dynamic difficulty adaptation (DDA) algorithms to keep players engaged and motivated [24, 87]. As shown in this and previous research [175], difficulty levels directly influence the players' affective states and consequently, their cognitive performance. When challenges and skills are balanced, players enter in a flow state [45] (see Fig. 2.4), also called optimal experience, that can help them to achieve a better cognitive performance [8]. In line with these findings, this research has demonstrated that players have a better WM performance when they are in an optimal affective state, feeling successful and overcoming difficult challenges.

Due to the effects of arousal and valence on cognitive skills, it is important to include affective state metrics in the adaptation engine in order to provide more effective cognitive training. As suggested by Mishra et al [98], closed-loop video games should include physiological and behavioural cues to achieve a more personalised cognitive training. This is especially relevant for game-based cognitive training since video games often involve emotional processes [24] and the pace of progression strongly depends on the player's skills. The adaptation engine implemented in *Memory Break* used affect and performance metrics for adaptation and decision making. Although the affect recognition system did not perform well, making wrong adaptation decisions, the performance-based decision layer corrected some of these erroneous decisions. These findings are in line with the research by Bontchev and Vassileva [25] who showed that adaptation accuracy can significantly improve when combining performance and affective metrics. Even though their research focused on computer games for educational purposes, their findings are reasonable as the adaptation engine would have a better picture of the player's responses. Whilst the performance decision layer keeps track of the player's progression, the affect decision layer appraises the player's affective responses, which can provides insights about how the player reacts to new challenges. This is important in gaming experiences since players can perceive challenges differently depending on their motivation [75, 58]. Since flow strongly depends on motivation and affective states [24], only performance-based adaptation is not enough to provide a tailored WM training that sustain players in an optimal affective state. This is supported by Lazarus' cognitive-motivational theory [83], which states that the subjects' motivation and goals in a situation must be known to understand their reactions.

The adaptive version of *Memory Break* in Study 3 had positive effects in participants' WM performance. However, only participants with low WM playing the adaptive version significantly improved their WM performance within *Memory Break* as well as their AOST score. These findings indicate that difficulty adaptation can be more beneficial for subjects with low WM capacity.

The design and implementation of the adaptation engine presented some challenges and limitations worth discussing. The most important limitation was the poor performance of the affect recognition system, which had a classification accuracy below the chance level of 50%. This could be explained by the novelty effect [123] and lower levels of self-reported arousal and valence of participants in Study 3 compared to those in Study 2. Another reason is the subject-independent approach used in the affect recognition system, which caused great accuracy variability between subjects. According to Parsons and Reinebold [119], subject-independent methods are harder to create than subject-dependent methods as they have to work for all subjects. Although subject-dependent methods are more reliable because they are exclusively created for each subject, they are also more expensive as they have to be calibrated for each subject. Due to the intensification of affective states in VR, developing automatic affect recognition systems in this medium can add some difficulty to this complex problem. Since VR is a relatively new medium, there can be important novelty effects on some participants.

The misclassification of arousal and valence involved an important challenge about how to deal with them. The buffers created to keep a time series of the previous predictions smoothed the impact of these misclassifications. Adaptation decisions were only taken if the engine was certain about the detected affective state (buffer mean $>.05$). Afergan et al [3] took a similar approach, averaging the previous 8 classification predictions to calculate a confidence value to make adaptation decisions. Since arousal and valence are continuous dimensions [138], it is important to keep track of how subjects' emotions progress in time. In addition, the creation of static thresholds is not a good solution to divide data for classification that is likely to have a great variability between subjects. The static threshold set up in Study 2 to divide arousal and valence into 2 classes for classification were problematic since not all subjects in studies 2 and 3 used the same range of values with the Affective Slider. This could be solved normalising participants' self-reported affective states.

Finally, the performance-based decision layer could be substantially improved, adding more meaningful features such as reaction times or errors made. The static threshold to differentiate good and bad scores was again problematic due to high variability between subjects in skills and expectations on what a good performance is. Montani et al [103] suggested a continuous calibration of the game's difficulty level depending on the player's current performance to account for performance variability. Therefore, it is important to account for the variance of subjects' self-reported variables, preferences and motivations. Further research needs to be done to address this important problem in affective computing research.

7.2 Contributions

The research presented in this PhD thesis revolves around three elements or variables: affective states, WM and VR. Previous research has demonstrated that adaptation in games for cognitive training can lead to cognitive improvements. This research explored the employment of HMD-based VR video games for WM training as well as the effects of immersion and affective states on WM performance. An adaptation engine was designed, implemented and tested to dynamically adjust the difficulty of the game according to the player's affective states and performance. The results demonstrated that affect and performance-based adaptation in VR games can have positive effects on the player's WM performance

This research contributes to the areas of affective gaming and game-based cognitive training in five ways. First, **high levels of immersion often reported in VR have a positive impact on player's WM performance and intensify their affective states**. Second, **positive affective states help subjects to achieve a better WM performance**, although arousal can have detrimental effects if it is too high. Third, **a new methodology for affect recognition in gesture-based VR gaming** using physiological and behavioural signals. This methodology proposed and demonstrated the feasibility of employing the HMD to measure the player's affective states. Fourth, **a novel adaptation engine composed by affect and performance metrics** to dynamically adjust the difficulty of the game. This adaptation engine aims to keep the player in an optimal affective state to improve the WM performance. Last, **difficulty adaptation can have positive effects on WM performance, being particularly pronounced on subjects with low WM capacity**. To the best of the researcher's knowledge, this is the first time that affect and performance metrics have been used for the adaptation of a game-based WM training program in VR.

The contributions of this PhD have demonstrated that VR is an effective medium for cognitive training as it improves subject's WM performance due to the high levels of immersion and engagement experienced. The findings reported also suggest that affect and performance metrics should be used to adapt the game in real-time to provide a tailored adaptation. Due to the effects of arousal and valence on subject's cognition, this research proposes the incorporation affective state metrics in the adaptive engine of a game-based cognitive training programs to improve its effectiveness.

Finally, this thesis suggests further research to investigate the effects of adaptive games in VR for cognitive training on subject's WM. Since adaptive games are more beneficial for subjects with low WM capacity, as shown in this thesis, further research should focus on subjects diagnosed with Dyslexia or ADHD as they suffer of low WM capacity [78].

7.3 Limitations

This section discusses some of the limitations of the research reported in this thesis. Suggestions are provided to address these limitations in future works.

7.3.1 Challenges in game type and interactions

The cognitive demands of video game playing can substantially change depending on the game genre played. According to Adams [2], there are more than 7 main genres (i.e.: Action, Adventure, Strategy, Sports, etc...), which are defined by its most common challenge and not by its content or medium of play. As Cohen, Green and Bavelier demonstrated [43], not all types of video games are beneficial for our visual attention as they challenge different parts of our cognition. Action video games, for instance, have been thoroughly studied for their positive effects on visuospatial cognition [43].

Memory Break, is a game designed to engage and immerse players while challenging their verbal WM. Although it demands a very simple interaction, different results could be obtained depending on the game played and its interaction. Other genres such as action video games could have different effects on players' cognition due to variations in the cognitive load. For example, if the game would required a convoluted sequence of actions, the player experience and cognitive performance might change. Future work should use appropriate games that do not overload the player's cognitive skills.

On the other hand, the difficulties reported with the interaction gestures used in *Memory Break* could have affected the results of this research. Before the development of a video game, a designer should consider the capabilities and constrains of the technology used, especially if it is novel. Designing hand gestures for game interactions can be challenging unless they are natural gestures that players are used to. Video games should use controllers that are easy to understand and use for all type of users.

7.3.2 Challenges in affect recognition

This research has discussed relevant challenges in the detection and recognition of players' affective states (see section 7.1.1). The results of Study 1 revealed that GSR sensors are not appropriate for affective state detection in video game playing as it is prone to noise when motion is involved [105, 154]. Participants in this study found the physiological sensors very invasive due to the electrodes attached to their chest and hand (see Fig. 3.2). This problem was solved in studies 2 and 3 using wearable sensors. Caution must be taken when interpreting physiological signals for affect recognition, since they are not exclusively related to one affective state. For example, an increase in cardiac activity is related to a high emotional arousal, while low activity is related to information processing and engagement [139]. Since video games can produce both high arousal and attentional engagement, HR activity may not be a good measure of arousal in these scenarios [127].

Another important limitation of this research was the poor performance of arousal and valence classification algorithms. Various problems can be identified in the design and training of the machine learning algorithms. First, participants in Study 2 should have reported arousal and valence levels at each section of the game as in Study 3. This would have generated more data for training and avoided the section selection reported in section 5.1.4. Second, the data used to train the classification algorithms should have been divided two chunks: 60% for training and 40% for testing. This would have made the classification testing easier, before embarking on Study 3. Due to the small amount of data available, all of it was used for training. These two limitations were critical on the design and development of the affect detection system, contributing to its poor classification accuracy. This problem could have been solved using a subject-dependent approach, although more data would have been necessary to train the models for each participant. The last problem on affect recognition was the significant differences in self-reported levels of arousal and valence between Study 2 and Study 3. These individual differences could have been solved by normalising the reported arousal and valence. Likewise to the normalisation of physiological signals, participants should have self-reported their arousal and valence levels at the beginning of each session. By not doing this, the median of the reported arousal and valence in Study 2 had to be used to divide them into two classes for classification.

7.4 Future Work

Parts of this thesis have been published (see section 1.4) and cited by some researchers. The paper "*Measuring affective, physiological and behavioural differences in solo, competitive and collaborative games*", associated with chapter 3, was cited for its insights about the impact of collaboration and competition game playing on players' engagement and physiological responses [136, 118, 52]. The paper associated with chapter 4, titled "*Effects of valence and arousal on working memory performance in virtual reality gaming*", was cited by [26] in a review about the use of VR to explore the effectiveness of media messages in

Figure 7.1: The Quantified behaviour Test (QbTest) [129] analyse the head motion to assess attention deficits and hyperactivity.

social campaigns.

In this final section, a number of areas for future work are briefly described following the findings of this research. These present opportunities to explore and enhance the contributions reported in this thesis.

7.4.1 Evaluation on subjects with cognitive impairments

All participants that took part in this research were healthy subjects that were not diagnosed with any learning difficulties such as Dyslexia or Attention Deficits and Hyperactivity Disorder (ADHD). Since the effects of adaptation were stronger in participants with low WM capacity, it would be interesting to test the benefits of playing *Memory Break* on subjects diagnosed with these learning difficulties. These subjects are known to suffer from deficits in executive functions such as WM [78]. Rizzo et al [133] used a VR classroom to assess this technology in the study of attention deficits in children with ADHD. Using a HMD, head rotation and general motor movement assessed distraction and hyperactive behaviour. They concluded that VR technology could improve cognitive problem assessment as it allows a higher control of stimuli presentation and a more precise measure of responses. This suggests that VR could be a good medium not only for cognitive training but also for cognitive assessment, using the HMD's sensors. This approach has been already implemented in the Quantified behaviour Test (QbTest) [129], a Continuous Performance Task that measures the two core symptoms of ADHD (attention and hyperactivity) measuring subject's head motion (see Fig. 7.1). This data is used to estimate how focused the subject is on the given task.

It would also be interesting to explore whether performance-based adaptation on its own would produce the same effects on participants with low WM capacity. Like the research of Liu et al [87], two performance-based adaptation systems could be tested with and without affective metrics to assess its benefits in subjects diagnosed with ADHD.

7.4.2 The future of VR in cognitive and wellbeing research

The introduction of HMDs such as Oculus or HTC Vive has allowed researchers an easy and affordable way of using VR. This has increased the number of VR applications for health and wellbeing. Various companies [1] and researchers [23] are using VR games to treat binocular vision issues like amblyopia or strabismus, replacing eye patches. Researchers like Mel Slater are exploring the use of VR to transform the self. For example, one of his latest studies suggested that having a virtual out-of-body experience reduces fear of death [27].

This research has used standing VR instead of room-scale VR - where participants can physically move around the VR environment - due to technological restrictions when the research started in 2014. Future work in the research of VR for cognitive training and wellbeing should incorporate new technologies such as room-scale environments [114] or haptics [27] to explore their impact on cognitive skills.

7.5 Closing remarks

This thesis has largely focused on the development of WM training programs using adaptive video games in VR. One of the main interests of this research was to explore the use of HMD-based VR applications for cognitive training. However, many authors use the term VR to refer to 3D environments experienced through a PC screen. This research aims to highlight the important differences in self-reported levels of immersion between immersive VR (using HMDs or CAVEs) and non-immersive VR (using PC screens or touchscreens) [114]. Due to the effects of immersion on cognitive skills, it is important to distinguish the results of studies using each interaction mode in order to understand the effects of immersion.

Finally, it should be noted that developing a video game is not a trivial task. It is a very long process that requires the expertise of graphic designers, sound engineers, developers and interaction designers. *Memory Break* was not designed to be a visually stunning video game, but to immerse and motivate players as well as having full control over game. Due to the complex task of developing a video game, the researcher encourages the collaboration of experts of different disciplines to create high quality video games.

[1] `https://www.seevividly.com`

Appendix A

Study 1 Material

A.1 Ethics Approval

Queen Mary
University of London

Queen Mary, University of London
Room W117
Queen's Building
Queen Mary University of London
Mile End Road
London E1 4NS

Queen Mary Ethics of Research Committee
Hazel Covill
Research Ethics Administrator
Tel: +44 (0) 20 7882 7915
Email: h.covill@qmul.ac.uk

Dr Laurissa Tokarchuk
CS302, Peter Landin Building
Department of Electronic Engineering
Queen Mary University of London
Mile End
London

16th June 2015

To Whom It May Concern:

Re: QMERC1502a – Collaboration and competition in computer games.

I can confirm that Mr Daniel Arellano has completed a Research Ethics Questionnaire with regard to the above research.

The result of which was the conclusion that his proposed work does not present any ethical concerns: is extremely low risk; and thus does not require the scrutiny of the full Research Ethics Committee.

Yours faithfully

Ms Hazel Covill – QMERC Administrator

A.2 Pre-Experiment Questionnaire

Questionnaire 1. Pre-Experiment

Please fill up the questionnaire below

1. **ID**

   ____

2. **Age**

   ____

3. **Gender**

   Mark only one oval.

   - ◯ Male
   - ◯ Female

4. **Occupation**

   ____

5. **1- Do you play videogames frequently?**

   Mark only one oval.

   - ◯ Yes
   - ◯ No

6. **2- How many times a week do you play video games?**

   Mark only one oval.

   - ◯ None
   - ◯ 1-2
   - ◯ 3-4
   - ◯ More than 4

7. **3- How many hours have you played video games in the last month?**

   Mark only one oval.

   - ◯ None
   - ◯ 0 - 2 hours
   - ◯ 3 - 5 hours
   - ◯ 6 - 10 hours
   - ◯ 11 - 15 hours
   - ◯ More than 15 hours

8. **4- On what platform do you usually play?**
Tick all that apply.

- [ ] PC
- [ ] PlayStation
- [ ] Xbox 360
- [ ] Xbox Kinect
- [ ] Wii
- [ ] Handheld consoles(PlayStation Portable or Vita, Nintendo DS...)
- [ ] Mobile Phone (iPhone, Android, Windows Phone...)
- [ ] Other: ______________________

9. **5- What is your favourite kind of video game?**
Tick all that apply.

- [ ] Action (platforms, fighting...)
- [ ] Shooter (First-person shooting, on-rails shooting...)
- [ ] Adventure (Horror, graphic adventure, interactive movie...)
- [ ] RPG (Role-playing, fantasy, western...)
- [ ] Simulation (Construction, life or vehicle simulation...)
- [ ] Strategy (Real-time tactics, tower defense, wargame...)
- [ ] Sports (Racing, football, basketball, tennis...)
- [ ] Puzzles (Brain training, logic...)
- [ ] Other: ______________________

10. **6- How do you prefer to play?**
Tick all that apply.

- [ ] Alone
- [ ] With someone (co-operative)
- [ ] Against someone (competitive)
- [ ] Online

11. **How do you see/describe yourself? Answer for each statement below.**
``I see myself as someone who…''
Mark only one oval per row.

	Disagree Strongly (1)	Disagree (2)	Neither agree nor disagree (3)	Agree (4)	Agree Strongly (5)
is reserved	◯	◯	◯	◯	◯
is generaly trusting	◯	◯	◯	◯	◯
tends to be lazy	◯	◯	◯	◯	◯
is relaxed, handles stress well	◯	◯	◯	◯	◯
has a few artistic interests	◯	◯	◯	◯	◯
is outgoing, sociable	◯	◯	◯	◯	◯
tends to find fault with other	◯	◯	◯	◯	◯
does a thorough job	◯	◯	◯	◯	◯
gets nervous easily	◯	◯	◯	◯	◯
has an active imagination	◯	◯	◯	◯	◯

A.3 Post-Condition Questionnaire

Questionnaire 2. Post-Condition

Please fill the questionnaire below about the game condition you just played

1. **ID:**

2. **Play mode:**

3. **1) How fun did you find the game?**
Mark only one oval.

	1	2	3	4	5	
Little						Very

4. **2) How boring did you find the game?**
Mark only one oval.

	1	2	3	4	5	
Little						Very

5. **3) How hard did you find the game?**
Mark only one oval.

	1	2	3	4	5	
Little						Very

6. **4) How engaged were you with the game?**
Mark only one oval.

	1	2	3	4	5	
Little						Very

7. **5) How immersed were you in the game?**
Mark only one oval.

	1	2	3	4	5	
Little						Very

8. **6) To what extent did you feel motivated while playing?**
Mark only one oval.

	1	2	3	4	5	
Not at all	◯	◯	◯	◯	◯	A lot

9. **7) How much effort did you put into playing the game?**
Mark only one oval.

	1	2	3	4	5	
Very little	◯	◯	◯	◯	◯	A lot

10. **8) Did you play without thinking about how to play?**
i.e.: Without thinking how to throw the ball, how to rotate the camera...
Mark only one oval.

	1	2	3	4	5	
Not at all	◯	◯	◯	◯	◯	A lot

11. **9) How did you feel with your partner in terms of...? (if playing against the computer, with the computer)**
Mark only one oval per row.

	1 (Very little)	2	3	4	5 (A lot)
Enjoyment	◯	◯	◯	◯	◯
Boredom	◯	◯	◯	◯	◯
Fun	◯	◯	◯	◯	◯
Engagement	◯	◯	◯	◯	◯
Stress	◯	◯	◯	◯	◯
Frustration	◯	◯	◯	◯	◯

A.4 Post-Experiment Questionnaire

Questionnaire 3. Post-Experiment

Please fill the questionnaire below when the experiment has finishes

1. **ID:**

2. **1) Did you like the game overall?**
Mark only one oval.
   - Yes
   - No

3. **2) Did you find the game or the controller difficult to manage?**
Mark only one oval.
   - Yes
   - No

4. **3a) How much did you enjoy the solo mode?**
Mark only one oval.

	1	2	3	4	5	
Little						A lot

5. **3b) How much did you enjoy the co-operative mode?**
Mark only one oval.

	1	2	3	4	5	
Little						A lot

6. **3c) How much did you enjoy the competitive mode?**
Mark only one oval.

	1	2	3	4	5	
Little						A lot

7. **4a) How engaged with your partner did you feel in the solo mode?**
Mark only one oval.

	1	2	3	4	5	
Little						Very

8. **4b) How engaged with your partner did you feel in the co-operative mode?**
Mark only one oval.

	1	2	3	4	5	
Little	⬭	⬭	⬭	⬭	⬭	Very

9. **4c) How engaged with your partner did you feel in the competitive mode?**
Mark only one oval.

	1	2	3	4	5	
Little	⬭	⬭	⬭	⬭	⬭	Very

10. **5) Which play mode did you...**
Mark only one oval per row.

	Solo mode	Co-operative mode	Competitive mode
enjoy the most	⬭	⬭	⬭
find the most boring	⬭	⬭	⬭
find the most fun	⬭	⬭	⬭
find the most frustrating	⬭	⬭	⬭
find the most stressful	⬭	⬭	⬭

11. **Any further comments or suggestions?**

Powered by
Google Forms

Appendix B

Study 2 Material

B.1 Ethics Approval

Queen Mary
University of London

Queen Mary, University of London
Room W117
Queen's Building
Queen Mary University of London
Mile End Road
London E1 4NS

Queen Mary Ethics of Research Committee
Hazel Covill
Research Ethics Administrator
Tel: +44 (0) 20 7882 7915
Email: h.covill@qmul.ac.uk

c/o Dr Laurissa Tokarchuk
CS 302 - Department of Computer Science
Queen Mary University of London
Mile End Road
London

13th December 2016

To Whom It May Concern:

Re: QMREC1873a – Games 4 Brains

I can confirm that Mr Daniel Arellano has completed a Research Ethics Questionnaire with regard to the above research.

The result of which was the conclusion that his proposed work does not present any ethical concerns; is extremely low risk; and thus does not require the scrutiny of the full Research Ethics Committee.

Yours faithfully

Ms Hazel Covill – QMERC Administrator

Patron: Her Majesty the Queen
Incorporated by Royal Charter as Queen Mary and Westfield College, University of London

B.2 Pre-Experiment Questionnaire

Study 2.5 - Pre-Experiment Questionnaire

Affective gaming for cognitive training study - Pre-Experiment questionnaire.

*Required

1. **ID** *

2. **Age** *

3. **Gender** *
   Mark only one oval.
   - Male
   - Female

4. **Occupation**

5. **Have you ever been diagnosed with a specific learning difficulty, such as Dyslexia?** *
   Mark only one oval.
   - Yes
   - No
   - Other:

6. **How many hours have you played video games in the last week?** *
   Mark only one oval.
   - None
   - 0-2 hours
   - 3-5 hours
   - 5-7 hours
   - 7-10 hours
   - More than 10 hours

7. **On what platform do you usually play?** *
Tick all that apply.

- [ ] Smartphones (iPhone, Android, iPod...)
- [ ] Tablets (iPad, Samsung Galaxy Tab...)
- [ ] Computer / Laptop
- [ ] Wii
- [ ] Xbox Kinect
- [ ] Playstation or Xbox
- [ ] Handheld consoles (PlayStation Portable or Vita, Nintendo DS...)
- [ ] Virtual Reality (Oculus Rift, HTC Vive, Google Cardboard, Samsung Gear...)
- [ ] Other: ____________________

8. **What kind of video game genre do you usually play?** *
Tick all that apply.

- [ ] Action (Platforms, fighting...)
- [ ] FPS: First-Person Shooting (on-rails shooting...)
- [ ] Adventure (Horror, graphic adventure, interactive movie...)
- [ ] RPG: Role-Playing Game (Fantasy, western...Sometimes multiplayer)
- [ ] Simulation (Construction, life or vehicle simulation...)
- [ ] Strategy (Real-time tactics, tower defense, wargame...)
- [ ] Sports (Racing, football, basketball, tennis...)
- [ ] Puzzles (Logic, matching or ordering...)
- [ ] Cognitive games (Brain training, skills development)
- [ ] Casual Games (mini-games or quick to play such as inifite runners. I.e.: Angry birds, FarmVille, Candy Crush...). May fit in other genres
- [ ] Other: ____________________

9. **What is your motivation to play a video game?** *
Mark only one oval.

- ( ) Just for fun or to discover new experiences
- ( ) To kill the time when i am bored (i.e.: when commuting)
- ( ) To be the best player and beat others, or to get the highest score
- ( ) Other: ____________________

10. **Have you ever played a Virtual Reality game or experience (i.e. rollercoaster, 360 videos or similar)**
Mark only one oval.

- ( ) Yes
- ( ) No

B.3 Post-Condition Questionnaire

Study 2.5 - Post-Condition Questionnaire

To be fill after each level in each on the interaction modes

*Required

1. **What interaction mode have you just played? ***

   Mark only one oval.

   - Desktop
   - Virtual Reality

2. **ID ***

3. **How engaged were you with the game? ***

   Mark only one oval.

	0	1	2	3	4	
Not at all						Extremely

4. **What level have you just played? ***

   Mark only one oval.

   - 1
   - 2
   - 3

5. **Please indicate how you felt while playing the game for each of the items,on the following scale: (Please ask the experimenter if you do not understand any of the statements below) ***

   Mark only one oval per row.

	0 (Not at all)	1 (Slightly)	2 (Moderately)	3 (Fairly)	4 (Extremely)
I was interested in the game					
I felt successful					
I felt bored					
I found it impressive					
I forgot everything around me					
I felt frustrated					
I found it tiresome (tired or boring)					
I felt irritable					
I felt skilfull					
I felt completely absorbed					
I felt content (in a state of peaceful happiness.)					
I felt challenged					
I had to put a lot of effort into it					
I felt good					

6. **How immersed were you in the game?** *
Mark only one oval.

	0	1	2	3	4	
Not at all	◯	◯	◯	◯	◯	Extremely

7. **To what extent did you feel motivated while playing?** *
Mark only one oval.

	0	1	2	3	4	
Not at all	◯	◯	◯	◯	◯	Extremely

Affective Slider

Please response to the Affective Slider in the browser's tab next to this one.

B.4 Post-Experiment Questionnaire

Study 2.5 - Post-Session Questionnaire

Please respond this questionnaire about your overall experience playing the game on the platform you just played (either iPad or VR)

*Required

1. **ID** *

2. **What interaction mode have you just played?** *
Mark only one oval.
   - Desktop
   - Virtual Reality

3. **Did you like the game overall?** *
Mark only one oval.
   - Yes
   - No

4. **Did you find the interaction mode or the controller difficult to manage?** *
Mark only one oval.
   - Yes
   - No

5. **Overall, to what extent did you feel motivated to memorise and recall correctly the letters in the memory task?** *
Mark only one oval.

	0	1	2	3	4	
Not at all						Extremely

6. **Overall, to what extent did you feel motivated to get the best score in the game?** *
Mark only one oval.

	0	1	2	3	4	
Not at all						Extremely

7. **Overall, how engaged were you with the game in this interaction mode?** *
Mark only one oval.

	0	1	2	3	4	
Not at all						Extremely

8. **How immersed were you in the game in this interaction mode?** *
Mark only one oval.

	0	1	2	3	4	
Not at all	◯	◯	◯	◯	◯	Extremely

9. **How present did you feel playing the game in this interaction mode?** *
Mark only one oval.

	0	1	2	3	4	
Not at all	◯	◯	◯	◯	◯	Extremely

10. **What was your main motivation to play the game?** *
Mark only one oval.

- ◯ Just for fun or to discover new experiences
- ◯ To challenge myself and get my best score
- ◯ To be the best player on the game and get the highest score
- ◯ I had no motivation
- ◯ Other: ____________

11. **Which difficulty level did you ...** *
NOTE: The levels shown below (1, 2 and 3) do NOT refer to the level order you played but to each individual level. For example, if you played first in level 3 and you think this was the most challenging level, please select the option 3 and not 1. If you have any question, please ask the experimenter.
Mark only one oval per row.

	1	2	3
find more challenging (difficult)?	◯	◯	◯
find more boring?	◯	◯	◯
enjoy the most?	◯	◯	◯
feel more aroused or activated?	◯	◯	◯
feel more focused on the game?	◯	◯	◯

12. **Please indicate how you felt after you finished playing the game for each of the items,on the following scale: (Please ask the experimenter if you do not understand any of the statements below)** *
Mark only one oval per row.

	0 (Not at all)	1 (Slightly)	2 (Moderately)	3 (Fairly)	4 (Extremely)
I felt revived					
I felt bad					
I found it hard to get back to reality					
I felt guilty					
It felt like a victory					
I found it a waste of time					
I felt energised					
I felt satisfied					
I felt disoriented					
I felt exhausted					
I felt that I could have done more useful things					
I felt powerful					
I felt weary (tired)					
I felt regret					
I felt ashamed (embarrassed)					
I felt proud					
I had a sense that I had returned from a journey					

13. **Any further comments or suggestions?**

**** Please respond questions below ONLY if you have played in both interaction modes (Desktop & VR) ****

14. **What interaction mode did you...**
Mark only one oval per row.

	Desktop	Virtual Reality	Both equal
enjoy the most?			
find more immersive?			
find more challenging?			
feel more aroused or activated?			
feel more motivated to success?			
feel more focused?			

Appendix C

Adaptation Engine Material

C.1 Machine Learning Models

The following tables show the classification accuracies of all possible combinations fusing the sensor's features for affect recognition. Fourteen models were created to classify arousal, valence and 4 emotions.

	HMD (26 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	43.67	65.51	65.51
Logistic Reg.	31.03	60.91	55.17
SVM	55.17	65.51	65.51
NN	39.08	59.77	62.06
KNN	35.63	54.02	59.77
AdaBoost	43.67	64.36	58.62
Rand. Forests	45.97	58.62	62.06
J48	33.33	51.72	68.96

	LM (27 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	29.88	57.47	48.27
Logistic Reg.	27.58	59.77	48.27
SVM	42.52	58.62	57.47
NN	28.73	54.02	57.47
KNN	39.08	66.66	58.62
AdaBoost	29.88	50.57	48.27
Rand. Forests	43.67	58.62	60.91
J48	26.43	50.57	51.71

	HR (10 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	32.18	58.62	52.87
Logistic Reg.	44.82	55.17	64.36
SVM	43.67	58.62	50.57
NN	42.52	52.87	51.72
KNN	32.18	57.47	51.72
AdaBoost	32.18	57.47	52.87
Rand. Forests	37.93	52.87	51.72
J48	26.43	47.12	59.77

	MYO (7 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	49.42	59.77	63.21
Logistic Reg.	44.82	49.42	56.32
SVM	48.27	55.17	64.36
NN	41.37	56.32	60.91
KNN	31.03	49.42	57.47
AdaBoost	49.42	59.77	59.77
Rand. Forests	42.52	50.57	60.91
J48	47.12	58.62	67.81

	HMD+LM (53 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	35.63	65.51	55.17
Logistic Reg.	27.58	47.12	50.57
SVM	51.72	62.06	68.96
NN	33.33	50.57	63.21
KNN	36.78	55.17	67.81
AdaBoost	35.63	57.47	55.17
Rand. Forests	45.97	52.87	65.51
J48	35.63	54.02	64.36

	HR+MYO (17 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	39.08	58.62	57.47
Logistic Reg.	40.22	50.57	58.62
SVM	49.42	54.02	62.06
NN	31.03	47.12	54.02
KNN	32.18	47.12	58.62
AdaBoost	39.08	57.47	60.91
Rand. Forests	44.82	58.62	65.51
J48	28.73	48.27	52.87

	HMD+HR (36 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	42.52	66.66	67.81
Logistic Reg.	36.78	55.17	51.72
SVM	55.17	59.77	71.26
NN	39.08	56.32	56.32
KNN	35.63	62.06	54.02
AdaBoost	43.67	66.66	63.21
Rand. Forests	49.42	51.72	65.51
J48	34.48	56.32	66.66

	HMD+MYO (33 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	45.97	64.36	70.11
Logistic Reg.	29.88	62.06	57.47
SVM	55.17	64.36	68.96
NN	43.67	60.91	66.66
KNN	42.52	55.17	58.62
AdaBoost	44.82	63.21	62.06
Rand. Forests	50.57	58.62	70.11
J48	33.33	54.02	73.56

	HMD+LM+HR (63 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	37.93	67.81	55.17
Logistic Reg.	22.98	50.57	52.87
SVM	50.57	58.62	68.96
NN	34.48	49.42	58.62
KNN	39.08	55.17	66.66
AdaBoost	36.78	58.62	54.02
Rand. Forests	48.27	57.47	62.06
J48	36.78	44.82	65.51

	HMD+LM+MYO (60 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	37.93	65.51	58.62
Logistic Reg.	27.58	50.57	48.27
SVM	51.72	62.06	72.41
NN	36.78	57.47	55.17
KNN	47.12	67.81	70.11
AdaBoost	35.63	59.77	57.47
Rand. Forests	45.97	56.32	66.66
J48	37.93	49.42	71.26

	LM+HR (37 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	28.73	63.21	48.27
Logistic Reg.	22.98	51.72	49.42
SVM	44.82	60.91	55.17
NN	40.22	49.42	54.02
KNN	37.93	54.02	67.81
AdaBoost	28.73	55.17	49.42
Rand. Forests	45.97	56.32	52.87
J48	29.88	59.77	55.17

	LM+MYO (34 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	27.58	58.62	48.27
Logistic Reg.	22.98	54.02	43.67
SVM	43.67	59.77	58.62
NN	28.73	51.72	55.17
KNN	44.82	64.36	59.77
AdaBoost	27.58	52.87	48.27
Rand. Forests	45.97	56.32	63.21
J48	27.58	52.87	66.66

	HR+MYO+HMD (43 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	42.52	65.51	68.96
Logistic Reg.	32.18	50.57	52.87
SVM	56.32	63.21	73.56
NN	42.52	62.06	60.91
KNN	41.37	62.06	59.77
AdaBoost	40.22	64.36	59.77
Rand. Forests	51.72	56.32	68.96
J48	33.33	60.91	70.11

	HR+MYO+LM (44 Features)		
	4 Emotions	**Arousal**	**Valence**
Naïve Bayes	27.58	60.91	49.42
Logistic Reg.	27.58	51.72	52.87
SVM	50.57	60.91	63.21
NN	29.88	49.42	49.42
KNN	35.63	55.17	63.21
AdaBoost	27.58	59.77	49.42
Rand. Forests	47.12	57.47	62.06
J48	40.22	54.02	52.87

Appendix D

Study 3 Material

D.1 Ethics Approval

Queen Mary
University of London

Queen Mary, University of London
Room W117
Queen's Building
Queen Mary University of London
Mile End Road
London E1 4NS

Queen Mary Ethics of Research Committee
Hazel Covill
Research Ethics Administrator
Tel: +44 (0) 20 7882 7915
Email: h.covill@qmul.ac.uk

c/o Dr Laurissa Tokarchuk
CS 203, Peter Landin Building
School of Electronic Engineering
and Computer Science
Mile End
London

29th January 2018

To Whom It May Concern:

Re: QMREC2109a – Games 4 Brains.

I can confirm that Daniel Gábana Arellano has completed a Research Ethics Questionnaire with regard to the above research.

The result of which was the conclusion that his proposed work does not present any ethical concerns; is extremely low risk; and thus does not require the scrutiny of the full Research Ethics Committee.

Yours faithfully

Mr Jack Biddle – Research Approvals Advisor

Patron: Her Majesty the Queen
Incorporated by Royal Charter as Queen Mary
and Westfield College, University of London

D.2 Post-Condition Questionnaire

Study3 OK - Post-Game Questionnaire

To be fill after each level in each on the interaction modes

*Required

1. **Session** *

2. **ID** *

3. **How immersed were you in the game?** *
   Mark only one oval.

	0	1	2	3	4	
Not at all	⬭	⬭	⬭	⬭	⬭	Extremely

4. **How engaged were you with the game?** *
   Mark only one oval.

	0	1	2	3	4	
Not at all	⬭	⬭	⬭	⬭	⬭	Extremely

5. **Please indicate how you felt while playing the game for each of the items,on the following scale: (Please ask the experimenter if you do not understand any of the statements below)** *
   Mark only one oval per row.

	0 (Not at all)	1 (Slightly)	2 (Moderately)	3 (Fairly)	4 (Extremely)
I was interested in the game	⬭	⬭	⬭	⬭	⬭
I felt successful	⬭	⬭	⬭	⬭	⬭
I felt bored	⬭	⬭	⬭	⬭	⬭
I found it impressive	⬭	⬭	⬭	⬭	⬭
I forgot everything around me	⬭	⬭	⬭	⬭	⬭
I felt frustrated	⬭	⬭	⬭	⬭	⬭
I found it tiresome (tired or boring)	⬭	⬭	⬭	⬭	⬭
I felt irritable	⬭	⬭	⬭	⬭	⬭
I felt skilfull	⬭	⬭	⬭	⬭	⬭
I felt completely absorbed	⬭	⬭	⬭	⬭	⬭
I felt content (in a state of peaceful happiness.)	⬭	⬭	⬭	⬭	⬭
I felt challenged	⬭	⬭	⬭	⬭	⬭
I had to put a lot of effort into it	⬭	⬭	⬭	⬭	⬭
I felt good	⬭	⬭	⬭	⬭	⬭

6. **To what extent did you feel motivated to play?** *
Mark only one oval.

	0	1	2	3	4	
Not at all	◯	◯	◯	◯	◯	Extremely

7. **To what extent did you feel motion sickness while playing?** *
Mark only one oval.

	0	1	2	3	4	
Not at all	◯	◯	◯	◯	◯	Extremely

D.3 Post-Experiment Questionnaire

Study3 OK - Post-Session Questionnaire

*Required

1. **ID** *

2. **Session** *

3. **Did you like the game overall?** *
   Mark only one oval.

   ⬭ Yes
   ⬭ No

4. **Did you find the interaction difficult to control?** *
   Mark only one oval.

   ⬭ Yes
   ⬭ No

5. **Overall, to what extent did you feel motivated to memorise and recall correctly the letters in the memory task?** *
   Mark only one oval.

	0	1	2	3	4	
Not at all	⬭	⬭	⬭	⬭	⬭	Extremely

6. **Overall, to what extent did you feel motivated to get the best score in the game?** *
   Mark only one oval.

	0	1	2	3	4	
Not at all	⬭	⬭	⬭	⬭	⬭	Extremely

7. **Overall, how engaged were you with the game?** *
   Mark only one oval.

	0	1	2	3	4	
Not at all	⬭	⬭	⬭	⬭	⬭	Extremely

8. **How immersed were you in the game?** *
Mark only one oval.

	0	1	2	3	4	
Not at all	⬭	⬭	⬭	⬭	⬭	Extremely

9. **How present did you feel in the VR environment while playing the game?** *
Mark only one oval.

	0	1	2	3	4	
Not at all	⬭	⬭	⬭	⬭	⬭	Extremely

10. **What was your main motivation to play the game?** *
Mark only one oval.

- ⬭ Just for fun or to discover new experiences
- ⬭ To challenge myself and get my best score
- ⬭ To be the best player on the game and get the highest score
- ⬭ I had no motivation
- ⬭ Other: ____________________

11. **For each of the items, please indicate how you felt after you finished playing the game: (Please ask the experimenter if you do not understand any of the statements below)** *
Mark only one oval per row.

	0 (Not at all)	1 (Slightly)	2 (Moderately)	3 (Fairly)	4 (Extremely)
I felt revived	⬭	⬭	⬭	⬭	⬭
I felt bad	⬭	⬭	⬭	⬭	⬭
I found it hard to get back to reality	⬭	⬭	⬭	⬭	⬭
I felt guilty	⬭	⬭	⬭	⬭	⬭
It felt like a victory	⬭	⬭	⬭	⬭	⬭
I found it a waste of time	⬭	⬭	⬭	⬭	⬭
I felt energised	⬭	⬭	⬭	⬭	⬭
I felt satisfied	⬭	⬭	⬭	⬭	⬭
I felt disoriented	⬭	⬭	⬭	⬭	⬭
I felt exhausted	⬭	⬭	⬭	⬭	⬭
I felt that I could have done more useful things	⬭	⬭	⬭	⬭	⬭
I felt powerful	⬭	⬭	⬭	⬭	⬭
I felt weary (tired)	⬭	⬭	⬭	⬭	⬭
I felt regret	⬭	⬭	⬭	⬭	⬭
I felt ashamed (embarrassed)	⬭	⬭	⬭	⬭	⬭
I felt proud	⬭	⬭	⬭	⬭	⬭
I had a sense that I had returned from a journey	⬭	⬭	⬭	⬭	⬭

12. **Any further comments or suggestions?**

Powered by
Google Forms

Bibliography

[1] Emg data from the myo armband. https://support.getmyo.com/hc/en-us/articles/202536726-How-do-I-access-the-raw-EMG-data-from-the-Myo-armband-. Accessed: 2019-01-03.

[2] E. Adams. *Fundamentals of game design.* Pearson Education, 2014.

[3] D. Afergan, E. M. Peck, E. T. Solovey, A. Jenkins, S. W. Hincks, E. T. Brown, R. Chang, and R. J. Jacob. Dynamic difficulty using brain metrics of workload. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 3797–3806. ACM, 2014.

[4] J. Allaire, L. Bäckman, D. Balota, D. Bavelier, R. Bjork, and G. Bower. A consensus on the brain training industry from the scientific community. *Max planck institute for human development and stanford center on longevity. Available at http://longevity3. stanford. edu/blog/2014/10/15/the-consensus-on-the-brain-trainingindustry-from-the-scientific-community-2*, 2014.

[5] M. Altini and O. Amft. Hrv4training: Large-scale longitudinal training load analysis in unconstrained free-living settings using a smartphone application. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 2610–2613. IEEE, 2016.

[6] K. Amunts, O. Kedo, M. Kindler, P. Pieperhoff, H. Mohlberg, N. Shah, U. Habel, F. Schneider, and K. Zilles. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anatomy and embryology*, 210(5-6):343–352, 2005.

[7] J. A. Anguera, J. Boccanfuso, J. L. Rintoul, O. Al-Hashimi, F. Faraji, J. Janowich, E. Kong, Y. Larraburo, C. Rolle, E. Johnston, et al.

Video game training enhances cognitive control in older adults. *Nature*, 501(7465):97, 2013.

[8] J. A. Anguera and A. Gazzaley. Video games, cognitive exercises, and the enhancement of cognitive abilities. *Current Opinion in Behavioral Sciences*, 4:160–165, 2015.

[9] J. A. Anguera, F. M. Gunning, and P. A. Areán. Improving late life depression and cognitive control through the use of therapeutic video game technology: A proof-of-concept randomized trial. *Depression and anxiety*, 34(6):508–517, 2017.

[10] B. M. Appelhans and L. J. Luecken. Heart rate variability as an index of regulated emotional responding. *Review of general psychology*, 10(3):229, 2006.

[11] A. D. Argembeau, C. Comblain, and M. Linden. Affective valence and the self-reference effect: Influence of retrieval conditions. *British Journal of Psychology*, 96(4):457–466, 2005.

[12] F. G. Ashby, V. V. Valentin, and A. Turken. The effects of positive affect and arousal on working memory and executive attention. *Advances in consciousness research*, 44:245–288, 2002.

[13] S. Asteriadis, N. Shaker, K. Karpouzis, and G. N. Yannakakis. Towards player's affective and behavioral visual cues as drives to game adaptation. In *LREC Workshop on Multimodal Corpora for Machine Learning*, pages 6–9. Citeseer, 2012.

[14] D. Baas, A. Aleman, and R. S. Kahn. Lateralization of amygdala activation: a systematic review of functional neuroimaging studies. *Brain Research Reviews*, 45(2):96–103, 2004.

[15] R. D. . F. Bailes. Time series analysis as a method to examine acoustical influences on real-time perception of music. *Empirical Musicology Review*, 2010.

[16] R. M. Baños, C. Botella, M. Alcañiz, V. Liaño, B. Guerrero, and B. Rey. Immersion and emotion: their impact on the sense of presence. *CyberPsychology & Behavior*, 7(6):734–741, 2004.

[17] C. Becker-Asano, D. Sun, C. Scheel, B. Tuschen-Caffier, and B. Nebel. Analyzing for emotional arousal in hmd-based head movements during a virtual emergency. In *Proc. Workshop on Emotion and Computing*, 2013.

[18] E. Bekele, J. W. Wade, D. Bian, L. Zhang, Z. Zheng, A. Swanson, M. Sarkar, Z. Warren, and N. Sarkar. Multimodal interfaces and sensory fusion in vr for social interactions. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 14–24. Springer, 2014.

[19] K. A. Bennion, J. H. Ford, B. D. Murray, and E. A. Kensinger. Oversimplification in the study of emotional memory. *Journal of the International Neuropsychological Society*, 19(9):953–961, 2013.

[20] D. Bersak, G. McDarby, N. Augenblick, P. McDarby, D. McDonnell, B. McDonald, and R. Karkun. Intelligent biofeedback using an immersive competitive environment. In *Online Proc. Ubicomp 2001 Workshop on Designing Ubiquitous Computer Games*. Paper at the Designing Ubiquitous Computing Games Workshop at UbiComp, 2001.

[21] A. Betella and P. F. Verschure. The affective slider: A digital self-assessment scale for the measurement of human emotions. *PloS one*, 11(2):e0148037, 2016.

[22] D. Bian, J. W. Wade, A. Swanson, Z. Warren, and N. Sarkar. Physiology-based affect recognition during driving in virtual environment for autism intervention. In *PhyCS*, pages 137–145, 2015.

[23] J. Blaha and M. Gupta. Diplopia: A virtual reality game designed to help amblyopics. In *Virtual Reality (VR), 2014 iEEE*, pages 163–164. IEEE, 2014.

[24] B. Bontchev. Adaptation in affective video games: A literature review. *Cybernetics and Information Technologies*, 16(3):3–34, 2016.

[25] B. Bontchev and D. Vassileva. Affect-based adaptation of an applied video game for educational purposes. *Interactive Technology and Smart Education*, 14(1):31–49, 2017.

[26] A. Borawska, M. Borawski, and M. Łatuszyńska. The concept of virtual reality system to study the media message effectiveness of social campaigns. *Procedia Computer Science*, 126:1616–1626, 2018.

[27] P. Bourdin, I. Barberia, R. Oliva, and M. Slater. A virtual out-of-body experience reduces fear of death. *PloS one*, 12(1):e0169343, 2017.

[28] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[29] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny. The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, 45(4):624–634, 2009.

[30] T. Brosch, K. R. Scherer, D. M. Grandjean, and D. Sander. The impact of emotion on perception, attention, memory, and decision-making. *Swiss medical weekly*, 143:w13786, 2013.

[31] E. Brown and P. Cairns. A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1297–1300. ACM, 2004.

[32] B. B. . N. Bryan-Kinns. Designing collaborative musical experiences for broad audiences. In *Proc. ACM Conf. on Creativity & Cognition*, 2013.

[33] G. C. Burdea. Invited review: the synergy between virtual reality and robotics. *IEEE Transactions on Robotics and Automation*, 15(3):400–410, 1999.

[34] R. A. Calvo and S. D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, 1(1):18–37, 2010.

[35] G. Castellano, S. D. Villalba, and A. Camurri. Recognising human emotions from body movement and gesture dynamics. In *International Conference on Affective Computing and Intelligent Interaction*, pages 71–82. Springer, 2007.

[36] G. Chanel, J. M. Kivikangas, and N. Ravaja. Physiological compliance for social gaming analysis: Cooperative versus competitive playâŸĘ. *Interacting with Computers*, 24(4):306–316, 2012.

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

[38] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun. Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(6):1052–1063, 2011.

[39] L. Chittaro and F. Buttussi. Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE Trans. Vis. Comput. Graph.*, 21(4):529–538, 2015.

[40] T.-J. Chou and C.-C. Ting. The role of flow experience in cyber-game addiction. *CyberPsychology & Behavior*, 6(6):663–675, 2003.

[41] T. Christy and L. I. Kuncheva. Technological advancements in affective gaming: A historical survey. *GSTF Journal on Computing (JoC)*, 3(4), 2014.

[42] W. C. Clapp, M. T. Rubens, J. Sabharwal, and A. Gazzaley. Deficit in switching between functional brain networks underlies the impact of multitasking on working memory in older adults. *Proceedings of the National Academy of Sciences*, 108(17):7212–7217, 2011.

[43] J. E. Cohen, C. S. Green, and D. Bavelier. Training visual attention with video games: Not all games are created equal. *Computer games and adult learning*, pages 205–227, 2007.

[44] T. Constant, G. Levieux, A. Buendia, and S. Natkin. From objective to subjective difficulty evaluation in video games. In *IFIP Conference on Human-Computer Interaction*, pages 107–127. Springer, 2017.

[45] M. Csikszentmihalyi. Toward a psychology of optimal experience. In *Flow and the foundations of positive psychology*, pages 209–226. Springer, 2014.

[46] A. Curci, T. Lanciano, E. Soleti, and B. Rimé. Negative emotional experiences arouse rumination and affect working memory capacity. *Emotion*, 13(5):867, 2013.

[47] E. Dahlin, A. S. Neely, A. Larsson, L. Bäckman, and L. Nyberg. Transfer of learning after updating training mediated by the striatum. *Science*, 320(5882):1510–1512, 2008.

[48] R. T. Dean and W. T. Dunsmuir. Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior research methods*, 48(2):783–802, 2016.

[49] A. Diamond. Executive functions. *Annual review of psychology*, 64:135–168, 2013.

[50] S. Dovis, S. Van der Oord, R. W. Wiers, and P. J. Prins. Can motivation normalize working memory and task persistence in children with attention-deficit/hyperactivity disorder? the effects of money and computer-gaming. *Journal of abnormal child psychology*, 40(5):669–681, 2012.

[51] K. Duff, L. J. Beglinger, S. K. Schultz, D. J. Moser, R. J. McCaffrey, R. F. Haase, H. J. Westervelt, D. R. Langbehn, J. S. Paulsen, H. S. Group, et al. Practice effects in the prediction of long-term cognitive outcome in three patient samples: A novel prognostic index. *Archives of Clinical Neuropsychology*, 22(1):15–24, 2007.

[52] L. Dupont, M. Pallot, L. Morel, O. Christmann, V. Boly, and S. Richir. Exploring mixed-methods instruments for performance evaluation of immersive collaborative environments. *International Journal of Virtual Reality*, 17(2), 2017.

[53] J. D. Eastwood, D. Smilek, and P. M. Merikle. Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & psychophysics*, 63(6):1004–1013, 2001.

[54] R. El Kaliouby and P. Robinson. Real-time inference of complex mental states from facial expressions and head gestures. In *Real-time vision for human-computer interaction*, pages 181–200. Springer, 2005.

[55] S. Fairclough. A closed-loop perspective on symbiotic human-computer interaction. In *International Workshop on Symbiotic Interaction*, pages 57–67. Springer, 2015.

[56] J. L. Foster, Z. Shipstead, T. L. Harrison, K. L. Hicks, T. S. Redick, and R. W. Engle. Shortened complex span tasks can reliably measure working memory capacity. *Memory & cognition*, 43(2):226–236, 2015.

[57] K. Gilleade and J. Allanson. A toolkit for exploring affective interface adaptation in videogames. In *HCI International 2003*, pages 370–374, 2003.

[58] K. Gilleade, A. Dix, and J. Allanson. Affective videogames and modes of affective gaming: assist me, challenge me, emote me. *DiGRA 2005: Changing Views–Worlds in Play.*, 2005.

[59] P. M. Greenfield, C. Brannon, and D. Lohr. Two-dimensional representation of movement through three-dimensional space: The role of video game expertise. *Journal of applied developmental psychology*, 15(1):87–103, 1994.

[60] S. M. Grüsser, R. Thalemann, and M. D. Griffiths. Excessive computer game playing: evidence for addiction and aggression? *CyberPsychology & Behavior*, 10(2):290–292, 2006.

[61] H. Gunes and M. Pantic. Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners. In *International conference on intelligent virtual agents*, pages 371–377. Springer, 2010.

[62] H. Gunes and B. Schuller. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2):120–136, 2013.

[63] M. A. Hall. *Correlation-based Feature Subset Selection for Machine Learning.* PhD thesis, University of Waikato, Hamilton, New Zealand, 1998.

[64] J. Holmes, S. E. Gathercole, and D. L. Dunning. Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental science*, 12(4), 2009.

[65] E. Hudlicka. Affective computing for game design. In *Proceedings of the 4th Intl. North American Conference on Intelligent Games and Simulation*, pages 5–12, 2008.

[66] W. IJsselsteijn, W. Van Den Hoogen, C. Klimmt, Y. De Kort, C. Lindley, K. Mathiak, K. Poels, N. Ravaja, M. Turpeinen, and P. Vorderer. Measuring the experience of digital game enjoyment. In *Proceedings of Measuring Behavior*, pages 88–89, 2008.

[67] R. M. . K. Inkpen. Physiological indicators for the evaluation of co-located collaborative play. In *Proc. of ACM CSCW*, 2004.

[68] S. M. Jaeggi, M. Buschkuehl, J. Jonides, and P. Shah. Short-and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, 108(25):10081–10086, 2011.

[69] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton. Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9):641–661, 2008.

[70] M. J. Kane, A. R. Conway, T. K. Miura, and G. J. Colflesh. Working memory, attention control, and the n-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3):615, 2007.

[71] E. A. Kensinger and D. L. Schacter. Memory and emotion. *Handbook of emotions*, 3:601–617, 2008.

[72] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[73] E. J. Kim, K. Namkoong, T. Ku, and S. J. Kim. The relationship between online game addiction and aggression, self-control and narcissistic personality traits. *European psychiatry*, 23(3):212–218, 2008.

[74] K. Kim, M. Z. Rosenthal, D. J. Zielinski, and R. Brady. Effects of virtual environment platforms on emotional responses. *Computer methods and programs in biomedicine*, 113(3):882–893, 2014.

[75] J. M. Kivikangas, J. Kätsyri, S. Järvelä, and N. Ravaja. Gender differences in emotional responses to cooperative and competitive game play. *PloS one*, 9(7):e100318, 2014.

[76] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed. Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(4):1027–1038, 2011.

[77] T. Klingberg. Training and plasticity of working memory. *Trends in cognitive sciences*, 14(7):317–324, 2010.

[78] T. Klingberg, E. Fernell, P. J. Olesen, M. Johnson, P. Gustafsson, K. Dahlström, C. G. Gillberg, H. Forssberg, and H. Westerberg. Computerized training of working memory in children with adhd-a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(2):177–186, 2005.

[79] T. Klingberg, H. Forssberg, and H. Westerberg. Training of working memory in children with adhd. *Journal of clinical and experimental neuropsychology*, 24(6):781–791, 2002.

[80] A. Kołakowska. A review of emotion recognition methods based on keystroke dynamics and mouse movements. In *Human System Interaction (HSI), 2013 The 6th International Conference on*, pages 548–555. IEEE, 2013.

[81] I. Kotsia, S. Zafeiriou, and S. Fotopoulos. Affective gaming: A comprehensive survey. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 663–670. IEEE, 2013.

[82] H. M. Lau, J. H. Smit, T. M. Fleming, and H. Riper. Serious games for mental health: are they accessible, feasible, and effective? a systematic review and meta-analysis. *Frontiers in psychiatry*, 7:209, 2017.

[83] R. S. Lazarus and R. S. Lazarus. *Emotion and adaptation*. Oxford University Press on Demand, 1991.

[84] N. Lazzaro. Why we play games: Four keys to more emotion without story. *In Game Developers Conference*, 2004.

[85] D. T. Levin and D. J. Simons. Perceiving stability in a changing world: Combining shots and intergrating views in motion pictures and the real world. *Media Psychology*, 2(4):357–380, 2000.

[86] T. M. Libkuman, P. Nichols-Whitehead, J. Griffith, and R. Thomas. Source of arousal and memory for detail. *Memory & Cognition*, 27(1):166–190, 1999.

[87] C. Liu, P. Agrawal, N. Sarkar, and S. Chen. Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction*, 25(6):506–529, 2009.

[88] T. W. Malone and M. R. Lepper. Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning and instruction*, 3, 1987.

[89] R. L. Mandryk and M. S. Atkins. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International journal of human-computer studies*, 65(4):329–347, 2007.

[90] R. J. Matheis, M. T. Schultheis, L. A. Tiersky, J. DeLuca, S. R. Millis, and A. Rizzo. Is learning and memory different in a virtual environment? *The Clinical Neuropsychologist*, 21(1):146–161, 2007.

[91] B. G. Mathers. Students' perceptions of 'fun' suggest possibilities for literacy learning: You can be entertained and informed. *Reading Horizons*, 49(1):6, 2008.

[92] K. Mawjee, S. Woltering, N. Lai, H. Gotlieb, R. Kronitz, and R. Tannock. Working memory training in adhd: controlling for engagement, motivation, and expectancy of improvement (pilot study). *Journal of attention disorders*, 21(11):956–968, 2017.

[93] K. L. McClarty, A. Orr, P. M. Frey, R. P. Dolan, V. Vassileva, and A. McVay. A literature review of gaming in education. *Gaming in education*, pages 1–35, 2012.

[94] D. McColl, C. Jiang, and G. Nejat. Classifying a person's degree of accessibility from natural body language during social human–robot interactions. *IEEE transactions on cybernetics*, 47(2):524–538, 2017.

[95] A. McMahan. Immersion, engagement and presence. *The video game theory reader*, 67:86, 2003.

[96] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks. Physiological measures of presence in virtual environments. In *Proceedings of 4th International Workshop on Presence*, pages 21–23, 2001.

[97] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks Jr. Physiological measures of presence in stressful virtual environments. *Acm transactions on graphics (tog)*, 21(3):645–652, 2002.

[98] J. Mishra, J. A. Anguera, and A. Gazzaley. Video games for neurocognitive optimization. *Neuron*, 90(2):214–218, 2016.

[99] J. Mishra, D. Bavelier, and A. Gazzaley. How to assess gaming-induced benefits on attention and working memory. *Games for Health: Research, Development, and Clinical Applications*, 1(3):192–198, 2012.

[100] J. Mishra and A. Gazzaley. Closed-loop rehabilitation of age-related cognitive disorders. In *Seminars in neurology*, volume 34, page 584. NIH Public Access, 2014.

[101] J. Mishra and A. Gazzaley. Closed-loop cognition: the next frontier arrives. *Trends in cognitive sciences*, 19(5):242–243, 2015.

[102] A. Montagrin, T. Brosch, and D. Sander. Goal conduciveness as a key determinant of memory facilitation. *Emotion*, 13(4):622, 2013.

[103] V. Montani, M. De Filippo De Grazia, and M. Zorzi. A new adaptive videogame for training attention and executive functions: design principles and initial validation. *Frontiers in psychology*, 5:409, 2014.

[104] A. M. Moore, N. O. Rudig, and M. H. Ashcraft. Affect, motivation, working memory, and mathematics. *The Oxford handbook of numerical cognition*, 2014.

[105] E. Morgan, H. Gunes, and N. Bryan-Kinns. Using affective and behavioural sensors to explore aspects of collaborative music making. *International Journal of Human-Computer Studies*, 82:31–47, 2015.

[106] L. E. Nacke and C. A. Lindley. Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *arXiv preprint arXiv:1004.0248*, 2010.

[107] A. Negut, S.-A. Matu, F. A. Sava, and D. David. Virtual reality measures in neuropsychological assessment: a meta-analytic review. *The Clinical Neuropsychologist*, 30(2):165–184, 2016.

[108] J. Nijhar, N. Bianchi-Berthouze, and G. Boguslawski. Does movement recognition precision affect the player experience in exertion games? In *International Conference on Intelligent Technologies for interactive entertainment*, pages 73–82. Springer, 2011.

[109] P. Nolin, A. Stipanicic, M. Henry, Y. Lachapelle, D. Lussier-Desrochers, P. Allain, et al. Clinicavr: Classroom-cpt: A virtual reality tool for assessing attention and inhibition in children and adolescents. *Computers in Human Behavior*, 59:327–333, 2016.

[110] A. C. Oei and M. D. Patterson. Enhancing perceptual and attentional skills requires common demands between the action video games and transfer tasks. *Frontiers in psychology*, 6:113, 2015.

[111] A. Öhman. The psychophysiology of emotion: An evolutionary-cognitive perspective. *Advances in psychophysiology*, 2(79):127, 1987.

[112] E. Olmos-Raya, J. Ferreira-Cavalcanti, M. Contero, M. Castellanos-Baena, I. Chicci-Giglioli, and M. Alcañiz. Mobile virtual reality as an educational platform: A pilot study on the impact of immersion and positive emotion induction in the learning process. In *Eurasia Journal of Mathematics Science and Technology Education*, volume 14, pages 2045–2057. Eurasia Publishing House, 2018.

[113] A. Ortony, T. J. Turner, and S. J. Antos. A puzzle about affect and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(4):725, 1983.

[114] É. Ouellet, B. Boller, N. Corriveau-Lecavalier, S. Cloutier, and S. Belleville. The virtual shop: A new immersive virtual reality environment and scenario for the assessment of everyday memory. *Journal of neuroscience methods*, 303:126–135, 2018.

[115] A. M. Owen, A. Hampshire, J. A. Grahn, R. Stenton, S. Dajani, A. S. Burns, R. J. Howard, and C. G. Ballard. Putting brain training to the test. *Nature*, 465(7299):775, 2010.

[116] R. J. Pagulayan, K. Keeker, D. Wixon, R. L. Romero, and T. Fuller. *User-centered design in games*. CRC Press Boca Raton, FL, 2002.

[117] F. Pallavicini, A. Ferrari, A. Zini, G. Garcea, A. Zanacchi, G. Barone, and F. Mantovani. What distinguishes a traditional gaming experience from one in virtual reality? an exploratory study. In *International Conference on Applied Human Factors and Ergonomics*, pages 225–231. Springer, 2017.

[118] M. Pallot, L. Dupont, O. Christmann, S. Richir, B. Vincent, and L. Morel. Ice breaking: Disentangling factors affecting the performance of immersive co-creation environments. In *VRICâĂŹ17 Proceedings of the Virtual Reality International Conference*, 2017.

[119] T. D. Parsons and J. L. Reinebold. Adaptive virtual environments for neuropsychological assessment in serious games. *IEEE Transactions on Consumer Electronics*, 58(2), 2012.

[120] E. A. Phelps. Human emotion and memory: interactions of the amygdala and hippocampal complex. *Current opinion in neurobiology*, 14(2):198–202, 2004.

[121] R. W. Picard et al. Affective computing. *Perceptual Computing Section, Media Laboratory, MIT*, 1995.

[122] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, 23(10):1175–1191, 2001.

[123] J. Poppenk, S. Köhler, and M. Moscovitch. Revisiting the novelty effect: When familiarity, not novelty, enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5):1321, 2010.

[124] P. J. Prins, S. Dovis, A. Ponsioen, E. Ten Brink, and S. Van der Oord. Does computerized working memory training with game elements enhance motivation and training efficacy in children with adhd? *Cyberpsychology, behavior, and social networking*, 14(3):115–122, 2011.

[125] A. K. Przybylski, C. S. Rigby, and R. M. Ryan. A motivational model of video game engagement. *Review of general psychology*, 14(2):154, 2010.

[126] N. Ravaja. Contributions of psychophysiology to media research: Review and recommendations. *Media Psychology*, 6(2):193–235, 2004.

[127] N. Ravaja, J. Laarni, T. Saari, K. Kallinen, and M. Salminen. Phasic psychophysiological responses to video game events: New criterion variables for game design. In *Proceedings of HCI*, 2005.

[128] N. Ravaja, T. Saari, M. Salminen, J. Laarni, and K. Kallinen. Phasic emotional reactions to video game events: A psychophysiological investigation. *Media Psychology*, 8(4):343–367, 2006.

[129] V. Reh, M. Schmidt, L. Lam, B. G. Schimmelmann, J. Hebebrand, W. Rief, and H. Christiansen. Behavioral assessment of core adhd symptoms using the qbtest. *Journal of attention disorders*, 19(12):1034–1045, 2015.

[130] L. P. Rieber. Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology*, 83(3):318, 1991.

[131] G. Riva, F. Davide, and W. A. IJsselsteijn. *Being there: Concepts, effects and measurements of user presence in synthetic environments.* Ios Press, 2003.

[132] G. Riva, F. Mantovani, C. S. Capideville, A. Preziosa, F. Morganti, D. Villani, A. Gaggioli, C. Botella, and M. Alcañiz. Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, 10(1):45–56, 2007.

[133] A. A. Rizzo, J. G. Buckwalter, T. Bowerly, C. Van Der Zaag, L. Humphrey, U. Neumann, C. Chua, C. Kyriakakis, A. Van Rooyen, and D. Sisemore. The virtual classroom: a virtual reality environment for the assessment and rehabilitation of attention deficits. *CyberPsychology & Behavior*, 3(3):483–499, 2000.

[134] C. E. Rolle, B. Voytek, and A. Gazzaley. Exploring the potential of the ipad and xbox kinect for cognitive science research. *Games for health journal*, 4(3):221–224, 2015.

[135] P. J. Rosa, C. Sousa, B. Faustino, F. Feiteira, J. Oliveira, P. Lopes, P. Gamito, and D. Morais. The effect of virtual reality-based serious games in cognitive interventions: a meta-analysis study. In *Proceedings of the 4th Workshop on ICTs for improving Patients Rehabilitation Research Techniques*, pages 113–116. ACM, 2016.

[136] C. Roth and H. Koenitz. Evaluating the user experience of interactive digital narrative. In *Proceedings of the 1st International Workshop on Multimedia Alternate Realities*, pages 31–36. ACM, 2016.

[137] A. Rovira, D. Swapp, B. Spanlang, and M. Slater. The use of virtual reality in the study of people's responses to violent incidents. *Frontiers in Behavioral Neuroscience*, 3:59, 2009.

[138] J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.

[139] T. Saari, N. Ravaja, J. Laarni, and M. Turpeinen. Emotional regulation system for emotionally adapted games. In *FuturePlay 2005 conference. East Lansing, Michigan State University, USA. October 13-15 2005*, 2005.

[140] J. L. Sabourin and J. C. Lester. Affect and engagement in game-basedlearning environments. *IEEE Transactions on Affective Computing*, 5(1):45–56, 2014.

[141] L. Satyen and K. Ohtsuka. *Strategies to Develop Dual Attention Skills Through Video Game Training.* PhD thesis, Ashgate, 2001.

[142] N. Savva and N. Bianchi-Berthouze. Automatic recognition of affective body movement in a video game scenario. In *International Conference on Intelligent Technologies for interactive entertainment*, pages 149–159. Springer, 2011.

[143] N. Savva, A. Scarinzi, and N. Bianchi-Berthouze. Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience. *IEEE Transactions on Computational Intelligence and AI in games*, 4(3):199–212, 2012.

[144] B. J. Schmeichel, R. N. Volokhov, and H. A. Demaree. Working memory capacity and the self-regulation of emotional expression and experience. *Journal of personality and social psychology*, 95(6):1526, 2008.

[145] M. A. Schmuckler. What is ecological validity? a dimensional analysis. *Infancy*, 2(4):419–436, 2001.

[146] D. J. Seyler, E. P. Kirk, and M. H. Ashcraft. Elementary subtraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6):1339, 2003.

[147] Z. Shipstead, T. S. Redick, and R. W. Engle. Is working memory training effective? *Psychological bulletin*, 138(4):628, 2012.

[148] V. J. Shute, M. Ventura, and F. Ke. The power of play: The effects of portal 2 and lumosity on cognitive and noncognitive skills. *Computers & education*, 80:58–67, 2015.

[149] M. Slater, C. Guger, G. Edlinger, R. Leeb, G. Pfurtscheller, A. Antley, M. Garau, A. Brogni, and D. Friedman. Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 15(5):553–569, 2006.

[150] T. Sonne and M. M. Jensen. Chillfish: A respiration game for children with adhd. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 271–278. ACM, 2016.

[151] C. R. Souza. The accord .net framework. *São Carlos, Brazil: http://accord-framework. net*, 2014.

[152] J. Storbeck and G. L. Clore. With sadness comes accuracy; with happiness, false memory: Mood and the false memory effect. *Psychological Science*, 16(10):785–791, 2005.

[153] M. Suriya-Prakash, G. John-Preetham, and R. Sharma. Is heart rate variability related to cognitive performance in visuospatial working memory? *PeerJ PrePrints*, 3:e1377v1, 2015.

[154] J. Sykes and S. Brown. Affective gaming: measuring emotion through the gamepad. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 732–733. ACM, 2003.

[155] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel. Influence of mental stress on heart rate and heart rate variability. In *4th European conference of the international federation for medical and biological engineering*, pages 1366–1369. Springer, 2009.

[156] J. Tao and T. Tan. Affective computing: A review. In *International Conference on Affective computing and intelligent interaction*, pages 981–995. Springer, 2005.

[157] J. G. Taylor and N. F. Fragopanagos. The interaction of attention and emotion. *Neural networks*, 18(4):353–369, 2005.

[158] T. Tijs, D. Brokken, and W. IJsselsteijn. Creating an emotionally adaptive game. In *International Conference on Entertainment Computing*, pages 122–133. Springer, 2008.

[159] P. Toril, J. M. Reales, and S. Ballesteros. Video game training enhances cognition of older adults: a meta-analytic study. *Psychology and aging*, 29(3):706, 2014.

[160] L. M. Trick, F. Jaspers-Fayer, and N. Sethi. Multiple-object tracking in children: The "catch the spies" task. *Cognitive Development*, 20(3):373–387, 2005.

[161] G. Turpin. Effects of stimulus intensity on autonomic responding: The problem of differentiating orienting and defense reflexes. *Psychophysiology*, 23(1):1–14, 1986.

[162] N. Unsworth, R. P. Heitz, J. C. Schrock, and R. W. Engle. An automated version of the operation span task. *Behavior research methods*, 37(3):498–505, 2005.

[163] N. Vaughan, V. N. Dubey, T. W. Wainwright, and R. G. Middleton. A review of virtual reality based training simulators for orthopaedic surgery. *Medical engineering & physics*, 38(2):59–71, 2016.

[164] D. Villani, M. Lucchetta, A. Preziosa, and G. Riva. The role of interactive media features on the affective response: a virtual reality study. *International Journal on Human Computer Interaction*, 1(5):1–21, 2009.

[165] H. A. Wadlinger and D. M. Isaacowitz. Fixing our focus: Training attention to regulate emotion. *Personality and Social Psychology Review*, 15(1):75–102, 2011.

[166] N. Weinstein, A. K. Przybylski, and R. M. Ryan. Can nature make us more caring? effects of immersion in nature on intrinsic aspirations and generosity. *Personality and Social Psychology Bulletin*, 35(10):1315–1329, 2009.

[167] K. C. Welch, U. Lahiri, C. Liu, R. Weller, N. Sarkar, and Z. Warren. An affect-sensitive social interaction paradigm utilizing virtual reality environments for autism intervention. In *International Conference on Human-Computer Interaction*, pages 703–712. Springer, 2009.

[168] B. T. West, K. B. Welch, and A. T. Galecki. *Linear mixed models: a practical guide using statistical software*. Chapman and Hall/CRC, 2014.

[169] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[170] D. L. Woods, M. M. Kishiyama, E. W. Yund, T. J. Herron, B. Edwards, O. Poliva, R. F. Hink, and B. Reed. Improving digit span assessment of short-term verbal memory. *Journal of clinical and experimental neuropsychology*, 33(1):101–111, 2011.

[171] D. Wu, C. G. Courtney, B. J. Lance, S. S. Narayanan, M. E. Dawson, K. S. Oie, and T. D. Parsons. Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task. *IEEE Transactions on Affective Computing*, 1(2):109–118, 2010.

[172] G. N. Yannakakis and J. Hallam. Real-time game adaptation for optimizing player satisfaction. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(2):121–133, 2009.

[173] Y.-c. Yeh, G.-J. Lai, C. F. Lin, C.-W. Lin, and H.-C. Sun. How stress influences creativity in game-based situations: Analysis of stress hormones, negative emotions, and working memory. *Computers & Education*, 81:143–153, 2015.

[174] R. M. Yerkes and J. D. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology*, 18(5):459–482, 1908.

[175] L. Zhang, J. Wade, D. Bian, J. Fan, A. Swanson, A. Weitlauf, Z. Warren, and N. Sarkar. Multimodal fusion for cognitive load measurement in an adaptive virtual reality driving task for autism intervention. In *International Conference on Universal Access in Human-Computer Interaction*, pages 709–720. Springer, 2015.