Combined Head-Mounted Display Virtual Reality and Electroencephalography in the Study of Working Memory

Mr Adam Durnin, BSc (Hons), MSc

Thesis submitted for the Degree of Doctor of Philosophy

The University of Hull and the University of York

Hull York Medical School

April 2024

Abstract

Modern head-mounted display virtual reality (HMD-VR) is used to present immersive virtual environments which reduce external distractions during research. Working memory load (WML) is commonly measured using electroencephalography (EEG), which non-invasively records voltage potential difference responses at the scalp. Having combined HMD-VR and EEG is potentially advantageous for neurophysiological studies of WML as it would allow for the objective measurement of WML in fully controllable and immersive virtual environments. This thesis aimed to investigate the combined use of HMD-VR and EEG in the study of working memory.

A systematic review of WML comparisons between HMD-VR and alternative displays using neurophysiological measures found that the use of HMD-VR has a variable effect on WML relative to screen-based and non-virtual reality presentations. The effect on WML was dependent on HMD-VR configuration and task, but WML predominantly did not differ or was lower in HMD-VR. High-specification HMD-VR and EEG were successfully combined to acquire event related potentials in response to visually and auditorily presented questions in a working memory arithmetic addition task. A follow-up of the arithmetic study compared EEG data preprocessing steps (highpass filtering, lowpass/notch filtering, eye-based artifact removal) to minimise HMD-VR-related artifacts. A spatial navigation study comparing WML between high-specification HMD-VR and desktop-based virtual reality presentations during a learning and recall maze task found that the ratio between theta and alpha frequency band activity did not differ between displays. The main limitation identified is that cybersickness symptoms increased during HMD-VR conditions in the arithmetical and spatial navigation tasks.

Taking the results together, it is found that high-specification HMD-VR was successfully combined with EEG to acquire event related potential and frequency responses. High-specification HMD-VR did not increase WML relative to screen-based virtual reality, indicating its potential utility in research and real-world applications.

Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Term
ANOVA	Analysis of Variance
AR	Augmented Reality
BCI	Brain Computer Interface
BOLD	Blood-Oxygen Level Dependent
CAVE	Cave Automatic Virtual Environment
CLT	Cognitive Load Theory
DB-HMD-VR	Desktop-Based Head-mounted Display Virtual Reality
DB-VR	Desktop-Based Virtual Reality
df	Degrees of Freedom (in a statistical analysis)
DOF	Degrees of Freedom (of an HMD-VR device)
EBAR	Eye-Based Artifact Removal
EEG	Electroencephalography
ERD	Event Related Desynchronisation
ERP	Event Related Potential
ERS	Event Related Synchronisation
ERSP	Event-Related Spectral Perturbation
F	F Statistic
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional near-infrared spectroscopy
FOV	Field-Of-View
HMD	Head-Mounted Display
HMD-VR	Head-Mounted Display Virtual Reality
Hp	Highpass
HS-HMD-VR	High specification head-mounted display virtual reality
Hz	Hertz
ICA	Independent Component Analysis
IPD	Inter-Pupillary Distance
L	Left
Lp	Lowpass
LPC	Late Positive Component

LPP	Late Positive Potential
M	Mean
M-Diff	Mean Difference
ms	Milliseconds
NASA-TLX	NASA Task Load Index
NF	Notch Filter
ns	Not Statistically Significant
p	P-Value
PSD	Power Spectral Density
R	Right
s	Seconds
SB-VR	Screen-Based Virtual Reality
SD	Standard Deviation
SEM	Standard Error of the Mean
SSQ	Simulator sickness questionnaire
SWC	Slow Wave Component
t	T-Value
TAR	Theta/Alpha Ratio
UoH	University of Hull
VE	Virtual environment
VPP	Vertex Positive Potential
VR	Virtual Reality
VSWM	Visuospatial Working Memory
WML	Working Memory Load
ε	Greenhouse-Giesser Epsilon
ηp2	Partial Eta Squared
μV	Microvolts
χ^2	Chi-Square Statistic

Acknowledgements

I would like to thank my supervisors Dr Aziz Asghar, Dr Tony Bateson and Professor Stewart Martin for the knowledge, feedback, time, and opportunities they have given me over the course of this PhD project. I will always be grateful for their guidance in helping me to become the researcher I am, and I will carry on their lessons as I continue my career. Special thanks to my primary supervisor, Aziz, for his unwavering support of me from the day I arrived in Hull, during the uncertain times of the COVID-19 global pandemic, and to now as I reach the end of my project.

I would also like to thank my colleagues who I have worked alongside whilst in the Learning in Virtual Environments Laboratory. I would like to thank Dr Murat Aksoy for the guidance he has given me whilst learning EEG techniques and analysis methods, Mr Michael Johnstone & Dr Kate Yuen for their support in testing experiments and at the conferences we have attended, and Mr Matthew Barras & Mr Liam Booth for the technical expertise they offered. From setting up the laboratory, to testing out the theories we came up with, to just making the laboratory a friendly place to be, it has been wonderful working with you all.

I would like to thank my fellow HYMS and University of Hull PhD students who made me feel welcome at the university and introduced me to a range of research I would have never otherwise seen. I would also like to thank all of the people who gave up their time to volunteer for my experiments, who without I could not have completed my research.

On a personal level, I would like to thank my parents for their continued support and belief in me, my brother Matthew who has travelled halfway across the country more than once to help in my hour of need, my grandmother for checking in with me and making sure I’m doing well, and Jasper for simply being the best dog. I also thank my oldest friend Natalie and the friends I have made in Hull, each of whom have supported me during this project, and whose company I will cherish for years to come.

And most of all, I want to thank my dearest Helena for the unending love and support she has given me over the years. Without you, I may have never submitted this thesis, and I will be eternally grateful for everything you have done for me

Meur ras rag redyans.

Author’s Declaration

I confirm that this work is original and that if any passage(s) or diagram(s) have been copied from academic papers, books, the internet or any other sources these are clearly identified by the use of quotation marks and the reference(s) is fully cited. I certify that, other than where indicated, this is my own work and does not breach the regulations of HYMS, the University of Hull or the University of York regarding plagiarism or academic conduct in examinations. I have read the HYMS Code of Practice on Academic Misconduct, and state that this piece of work is my own and does not contain any unacknowledged work from any other sources. I confirm that any participant information obtained to produce this piece of work has been appropriately anonymised.

Chapter 1) The Utilisation of Head-Mounted Display Virtual Reality and Electroencephalography in Working Memory

1.1) Head-Mounted Display Virtual Reality as a Tool for Psychological Research

The term ‘virtual reality’ (VR) in psychological and neuroscience research broadly refers to any method of presenting a computer-generated virtual environment (VE) to participants [1,2]. Using VR methodologies in psychological studies was first proposed in the 1960s, and has evolved following the widespread adoption of computer technology in research during the 1970s [3–5]. Computer-based research offered several advantages over traditional verbal or pen-and-paper studies, primarily allowing for millisecond-accurate stimuli presentation and response measurement. Computers can also enhance experimental procedures in ways difficult for traditional methodologies, for example actively randomising elements of the experiment and providing automatic feedback to participants [4]. VR methodologies retain and expand upon the benefits of computerised presentation by allowing the use of fully controlled VEs, which can present participants with scenarios ranging from every-day occurrences to impossible events [1,6]. Today, using computer screens and VR technologies to present stimuli are main methodologies used in psychological research, and are standard in many neurophysiological recording configurations such as electroencephalography (EEG) [5,7,8].

Head-mounted display virtual reality (HMD-VR) refers to wearable goggle-like devices which facilitate naturalistic sensorimotor perception (‘high immersion’) by ‘placing users into’ stereoscopic VEs [9–11]. HMD-VR, also referred to as ‘immersive VR’ [12], builds upon the advantages of using VR in research by presenting VEs which can be ‘realistically’ explored and interacted with, for example turning one’s head to visually explore the VE, or replicating physical real-world actions to complete a task [13]. HMD-VR devices can also exclude external stimuli during use,

minimising confounding distractions [6] and producing high feelings of ‘presence’ [14–16], defined as the subjective feeling of ‘being transported’ or ’being there’ [6,9,10,17].

Experimental design within HMD-VR facilitates paradigms with high ecological validity [18], as research can be conducted in environments simulating real-world experiences, tasks and actions [19,20]. For example, research has been conducted within simulated classrooms during learning and social paradigms [21,22]. Moreover, paradigms or scenarios that are difficult or impossible to physically create due to the complexity, risk or costs involved can be conducted in HMD-VR. HMD-VR experiments emulating airport security checkpoints [23], crossing roads with oncoming traffic whilst distracted [24], and traversing balance beams suspended at high altitude [25] have been conducted. Individual factors and elements of the HMD-VR VE which are difficult to manipulate in real life can also be controlled, such as wind speed and direction in a marksmanship task [26], allowing consistency between experimental trials. Whilst similar paradigms can be conducted in other forms of VR, HMD-VR offers a more cost-effective and space-effective solution than other VR methodologies [27].

HMD-VR has been used in research since the 1990s [18,28] but was prevented from widespread adoption by the high costs associated with the devices [29]. However, the 2013 release of the first ‘modern’ HMD-VR device, the Oculus Rift DK1, has led to a range of affordable HMD-VR devices becoming accessible. The increased availability and technological advances has greatly expanded interest in the use of the HMD-VR in psychological and neuroscience research [5,20,29]. Modern HMD-VR devices are typically paired with internal or external sensors to offer 3-degrees of freedom (DOF) head-based rotational tracking, or full 6-DOF positional tracking in 3D space, allowing intuitive visual exploration and movement within a VE. Modern (i.e. post-2013) systems such as the Oculus Rift CV1 (Figure 1.1) also include accelerometer-enabled and motion-tracked handheld controllers and sensors to allow physical movements and object interactions to be represented within the VE. Object interactions utilising motion controllers also provide tactile feedback when

pressing a button or squeezing the controller to interact with or ‘hold’ a virtual object. The sensory feedback can be additionally enhanced by utilising vibrations within the controller to simulate touching or using an object [30]. However, lower-immersion configurations without motion controllers, such as those using a button on the side of the HMD-VR device [16], will be limited in the sensory feedback provided. Independent of input method, object interaction can be accompanied by visual feedback of the item moving or action being completed, either by the physical movements being represented by an avatar in the VE, or a preset animation being played. Many modern HMD-VR devices are also compatible with a range of peripheral input methods, for example steering wheels [31] and flight sticks [32], allowing for intuitive interaction with various VEs or virtual scenarios in research.

Figure 1.1: *Image of the Oculus Rift CV1 DB-HMD-VR display device* [33].

1.1.1) Types of Head-Mounted Display Virtual Reality

HMD-VR devices can be divided into several subcategories based on immersion, as determined by the perceptual facilitations and limitations imposed by a VR device. Aspects such as higher resolution screens, larger field-of-view (FOV), 6-DOF, intuitive interaction methods, and quality of the presented VE will increase immersiveness. Desktop-based HMD-VR (DB-HMD-VR) are displays that rely on secondary devices such as desktop computers to process the VE, and offer the highest levels of immersion. As minimal processing occurs within the HMD, DB-HMD-VR can support VEs requiring higher technical specifications when coupled with high-specification computers, providing more immersive experiences [6]. However, DB-HMD-VR is typically limited by movement-restricting tethering to the host device [34].

DB-HMD-VR can be further divided between consumer-grade DB-HMD-VR, and advanced ‘high-specification’ (HS-HMD-VR) models such as the HTC Vive Pro (Figure 1.2). HS-HMD-VR devices have improved technical specifications, such as higher resolution displays with larger FOV, allowing for VEs to be presented in higher clarity compared to standard DB-HMD-VR models.

Figure 1.2: *Image of the HTC Vive Pro HS-HMD-VR display device* [35].

Unlike DB-HMD-VR, standalone HMD-VR devices such as the Oculus Quest (Figure 1.3) are capable of presenting VEs without additional hardware. The processing of the VE is by performed in-device, facilitating easier field-deployment and lower overall costs [36]. However, due to the limited processing power of standalone HMD-VR devices, the VEs presented are of lower overall immersion.

Figure 1.3: *Image of the Oculus Quest standalone HMD-VR device* [37].

Smartphone HMD-VR are head-mounted adapters for smartphones running VR-emulating applications, such as the Google Cardboard (Figure 1.4). Smartphone HMD-VR is a cost-effective solution as many modern smartphones have some VR capability [16], but offers the reduced immersion due to low specifications, limited positional tracking, and restrictive input methods.

Figure 1.4: *Image of the Google Cardboard smartphone HMD-VR device* [38].

1.1.2) Other Virtual Reality Configurations

Traditionally, ‘VR’ within psychological research referred to Screen-Based Virtual Reality (SB-VR) systems, typically using desktop-based VR (DB-VR) configurations. When using DB-VR, the VE is presented on a standard computer screen and interacted with using a keyboard and mouse, but can use input methods tailored to the task presented. However, DB-VR is considered ‘low immersion’ [39] as it does not intuitively facilitate sensorimotor perception, instead resembling looking through a window into the VE.

Cave Automatic Virtual Environment (CAVE) systems [40,41] are ‘semi-immersive’ VR systems [39] which utilise projected imagery onto surrounding walls which, when viewed through 3D glasses, produces a stereoscopic VE. The participant stands centred within the CAVE system, and the VE appears to extend out into the distance or come towards the participant, effectively placing participants within the VE whilst retaining their own body. However, CAVE systems are limited by interactions with the VE, and cannot easily facilitate naturalistic physical object interaction compared to

HMD-VR VEs. For example, raising your hand to ‘grab’ an object can block the display or introduce visual disparity between the physical and virtual elements [42]. Specialist peripheral input methods or equipment configurations can facilitate accurate interactions for specific tasks, but can be overly complex and space-consuming [43].

1.2) Electroencephalography and Head-Mounted Display Virtual Reality

HMD-VR devices on the consumer market are generally designed to be ‘one-size-fits-all’, and use adjustable or elasticated straps to facilitate a range of head sizes. Many modern HMD-VR devices are also compatible with a range of replacement straps, which can expose different parts of the scalp depending on the model used. In research application, adjustable HMDs allows head-mounted neurophysiological recording equipment such as EEG to be used underneath the VR devices. It can be difficult to accommodate EEG in HMD-VR [20], and researchers sometimes modify HMD-VR devices to better facilitate EEG headcaps [44].

EEG is a non-invasive neurophysiological measure of electrical activity produced by the brain, captured though a series of electrodes placed on the scalp surface [45]. First used to record human brain activity in 1924 by Hans Berger and popularised worldwide by Adrian and Matthews in 1934, EEG is the predominant neurological method used today, and is commonly used in conjunction with behavioural and other physiological measures in cognitive research [45–48]. When presented with a stimulus, the neuronal response in associated brain regions results in a change in voltage that can be measured by EEG. The electrophysiological signal detectable by EEG results from the summed excitatory and inhibitory post synaptic potentials of the neurons underneath the recording site, which result in a positive and negative voltage change respectively [49].

The recordable electrical signal at the scalp surface generated by neuronal activity in the brain is very small (<100μV), as it must travel through meninges, skull, and scalp

before reaching the recording electrodes [45]. To acquire high-quality neurological data, modern EEG systems are comprised of a series of components that facilitate both the recording of the data and the execution of the experimental procedure.

The main component of the EEG system is the metallic recording electrodes sensitive to electrical signals placed against the scalp. EEG is commonly performed with a conductive gel or paste, which bridges the gap between the scalp and the electrode to reduce the impedance [50]. Electrodes are placed on standardised locations across the scalp, either individually or by using a pre-spaced headcap, using the international 10-20 system (Figure 1.5) based on percentage distances between electrodes from the scalp centre [45,51]. EEG also uses ground and reference electrodes to increase the signal-to-noise ratio. The ground electrode is a necessary part of the amplifier and recording systems, and is used to reduce noise from internal and external system sources. Reference electrodes are placed on or near the head where brain activity is not expected, such as on the mastoid bones or earlobes, to detect ambient electrical signals [45]. Subtracting the data captured by the reference electrodes from the EEG waveforms reduces external sources without compromising the EEG signal.

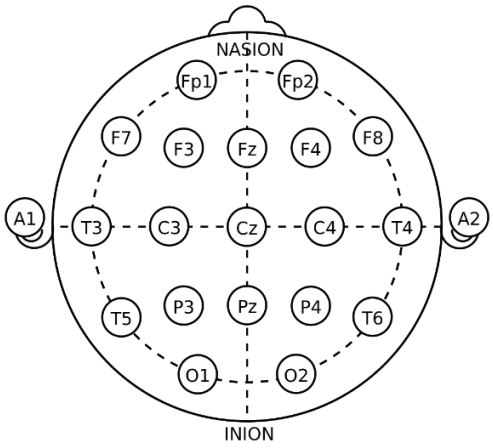


Figure 1.5: *Layout of a 19-channel EEG electrode configuration using the international 10-20 system with 2 reference electrodes (A*1 and A2) [52]. *The capital letters in the electrode designations refer to their location on the scalp: 'FP' refers to electrodes over the frontal pole (prefrontal), ''F' electrodes are over frontal lobe, 'C' electrodes are over the central region, 'T' electrodes are over the temporal lobe, 'P' electrodes are over the parietal lobe, 'O' electrodes are over the occipital lobe. Odd numbered electrodes are over the left hemisphere, even numbers are over the right hemisphere, and the 'z' electrodes are over the midline.*

The EEG electrodes/headcap connect to the acquisition device, which contains an input for the electrodes (the 'headbox'), an amplifier, an analog-to-digital converter, and a connection to the recording device (Figure 1.6). Signals recorded from the electrodes are enhanced by the amplifier, which increases the gain of each recorded signal by a factor of 1000-10000 [48,53]. Amplification also contributes to the online filtering of unwanted signals [45,48], as differential amplifiers remove common activity between the reference electrode and the recording electrode. The amplified signal undergoes analog-to-digital conversion [54], allowing recorded signals to be analysed in compatible digital formats. The converted EEG signal is outputted to the recording device, which can either be a specialised device or a computer running specialised software. The acquisition or recording device may also support a trigger system, which can mark important events in the data such as the onset of stimlui or button press responses.

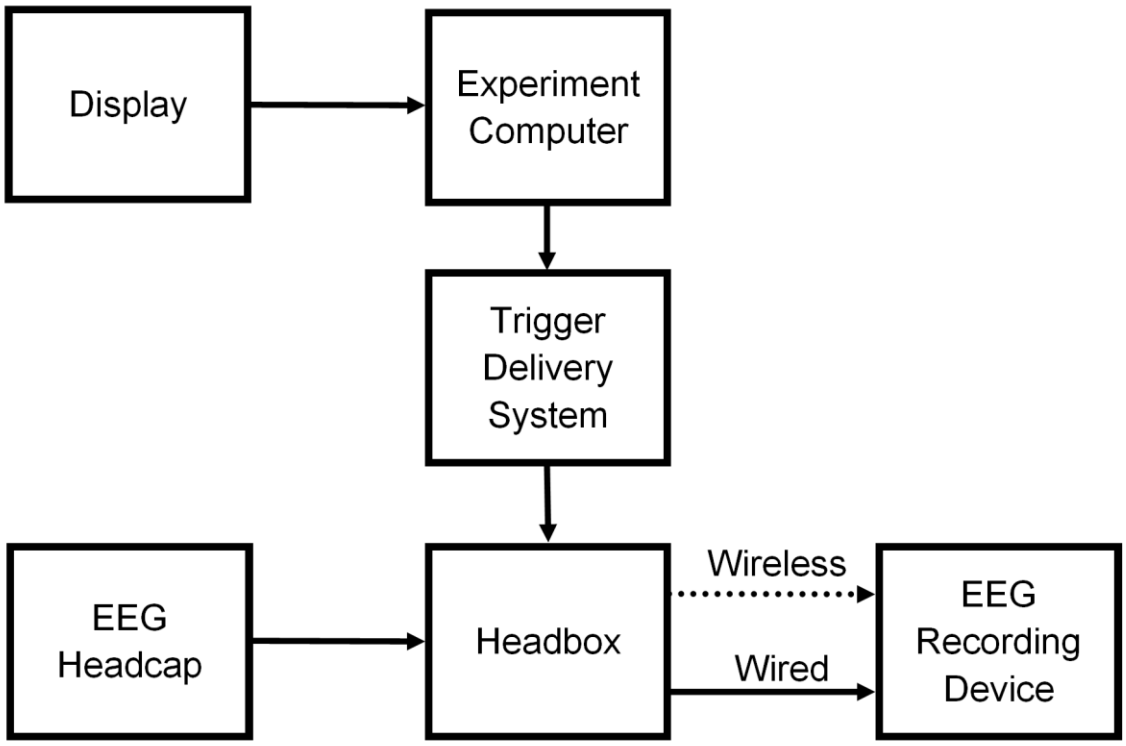


Figure 1.6: *A standard EEG configuration, including all major components used in a standard EEG experiment.*

Presentation devices are sometimes included in the EEG configuration, for example a VR device and supporting computer hardware. Dedicated presentation devices allow researchers to present stimuli with millisecond accuracy using specialised research software, such as PsychoPy [55]. The computers controlling the presentation device may additionally record behavioural data, and send triggers to the acquisition or recording device.

1.3) Combined Head-Mounted Display Virtual Reality and Electroencephalography in Research

Whilst the physical designs of the HMD-VR may not inherently exclude combined usage with EEG, it can be difficult to physically accommodate using both devices together [20]. EEG is sensitive to many sources of electrical interference, for example 50/60Hz line noise from the mains power, computer displays, and lighting [45]. It is therefore possible that HMD-VR, which can include electrically powered

screens in very close proximity to the recording electrodes, could introduce electrical noise to the EEG recording. To address this concern, initial research has been conducted examining the effect of different types of modern HMD-VR devices on EEG signal quality. Cattan et al. [36] demonstrated there was minimal-to-no effect on the quality of EEG signals below 36Hz when using smartphone HMD-VR. Moreover, Hertweck et al. [56] used time-frequency analysis to find that the DB-HMD-VR Oculus Rift and HS-HMD-VR Vive Pro introduced 50Hz line noise to the EEG recording, and an additional 90Hz noise when using the DB-HMD-VR. Line noise is also common in non-HMD-VR EEG data and can be easily removed during data preprocessing using digital filtering [57], minimising the unwanted signal.

There are several analysis methods that can be used in EEG experiments, which probe different aspects of the recorded signal and suit different paradigms used in research [45,58]. For example, event-related potentials (ERPs) examine changes in the time-locked amplitude and latency of voltage peaks within the ERP waveform [59]. A set amount of time surrounding each stimuli presentation, typically ~1s, is averaged over the collected trials (‘epochs’). The post-stimulus changes in amplitude are compared against the average amplitude of the pre-presentation baseline, where the brain should be at rest. The averaging process also increases the signal-to-noise ratio by reducing random noise to leave the targeted response. Comparisons can then be conducted between conditions or regions of the brain to examine neural responses to stimuli.

ERP methods have seen use in combined HMD-VR and EEG experiments across a range of HMD-VR subcategories. ERP studies have been conducted using non-modern pre-2013 HMD-VR [60–63] and smartphone HMD-VR [64]. Modern DB-HMD-VR devices such as the HTC Vive have been used in experiments acquiring ERP responses, for example during naturalistic reaching [65], and attention and memory-based tasks [66]. Moreover, a singular example of the HS-HMD-VR Vive Pro being utilised in ERP analysis could be found [67].

Whereas ERPs are locked to repeated events within a paradigm, time/frequency studies can involve longer durations as the changes at the power or amplitude of certain frequency bands of activity over time are compared. Typically, set frequency bands, for example theta (4-7Hz) and alpha (8-13Hz), are extracted and either averaged to find the power of a frequency band [68], compared against a baseline to examine relative changes in power [69], or visually inspected to find peaks of power in certain bands over time [70]. It can then be inferred what cognitive processes are represented by the changes in activity based on comparison between conditions and behavioural findings.

As with ERP studies, time-frequency measures have been used in research with several categories of HMD-VR. Time-frequency measures have been used to compare attention and memory in tasks presented with smartphone HMD-VR [16,36]. DB-HMD-VR headsets such as the HTC Vive have been used in EEG time-frequencies studies, including studying changes in frequency power over time during a physical spatial rotation task [71]. Time frequency analysis has also been used in a DB-HMD-VR experiment examining negative symptoms of motion sickness that can arise during VR usage called cybersickness [72]. There are also several examples of the HS-HMD-VR Vive Pro being used in time-frequency EEG experiments [56,73].

The examples of successfully combined HMD-VR and EEG are promising for the application of HMD-VR in research, demonstrating that the methods are not inherently incompatible. However, to date, several gaps in knowledge pertaining to combined HMD-VR and EEG methodologies can be identified. Certain subcategories of HMD-VR, particular HS-HMD-VR, are currently under-represented in the literature for ERP studies. Hyun & Lee's [67] example of HS-HMD-VR being used in an ERP study does not delve into the viability and feasibility of using such devices in wider research applications, nor are there examples of HS-HMD-VR being used in other tasks. Moreover, there is little-to-no guidance on how to preprocess EEG data captured in an HMD-VR experiment. Despite several EEG artifacts being identified during HMD-VR experiments [36], no dedicated examination of how to remove the

artifacts from data acquired during the combined use of HMD-VR and EEG has been conducted.

1.4) Working Memory and Cognitive Load Theory in Head-Mounted Display Virtual Reality

1.4.1) The Multicomponent Model of Working Memory

It has previously been found that increasing the immersiveness of a VR configuration, for example using a higher resolution DB-VR screen, can benefit response time and accuracy compared to low-immersion configurations [74,75]. However, the specific advantages of using HMD-VR over other types of VR in psychological and neuroscience research goes past a simple increase to technical specifications, as it produces a fundamentally different experience. Instead, the benefits of HMD-VR in research and application can be understood through the theories of working memory and cognitive load.

Baddeley and Hitch's multicomponent model of working memory is the leading theory of how information is actively held and processed online within the brain [76–79]. The multicomponent model builds upon and overcomes the limitations of Atkinson and Shiffrin's multi-store memory model [80], particularly expanding upon the short-term memory store located between sensory input and long-term memory (Figure 1.7a). Unlike the multi-store model, which assumes processing is independent of sensory modality, the multicomponent model attempts to explain how different types of sensory information is processed within working memory, and how this flow of information is controlled and combined.

Figure 1.7: *A simplified version of the multi-store model of memory demonstrating how each store interconnects. Information from each sensory modality is processed within the short term store, which can also encode information to and retrieve information from the long term store. The returning arrow on the short term store represents rehearsal.*

The multicomponent model describes four interconnected, but functionally distinct, subcomponents connected to sensory inputs and long term memory (figure 1.8b). The first two components explain how information in different modalities is processed from sensory input: the visuo-spatial sketchpad processes visual and spatial information; whilst the phonological loop processes auditory information and speech. These components route the processed information to the episodic buffer, which integrates the sensory inputs with information retrieved from long-term memory into a unified representation. The fourth subcomponent is the central executive, which acts as the overarching supervisory component controlling the working memory processes. Central executive functions include directing attention, managing the flow of information between the subcomponents and long-term memory, ordering the manipulation of information held in working memory, and making decisions based on the outcome of the information processing [81].

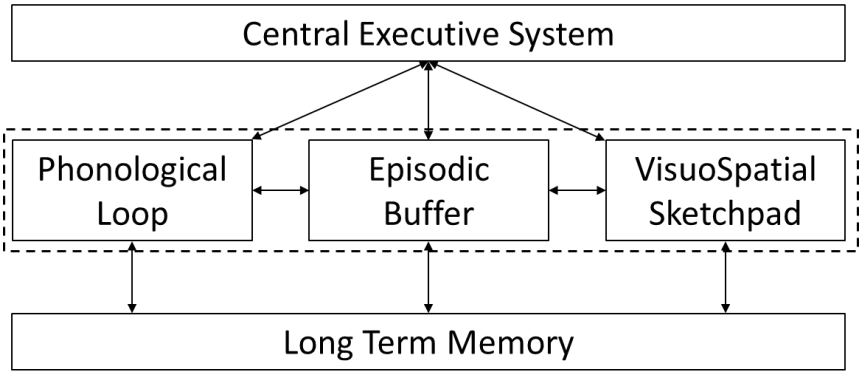


Figure 1.8: *A simplified version of the multicomponent model of working memory displaying how each subcomponent interacts. The central executive system can also direct the recall of memory and schema from the long term memory. Figure is adapted from Baddeley [78].*

The ability to distinguish between the processing of visual and auditory information, and describe how these may affect experienced working memory load (WML), is important for understanding the benefits and limitations of HMD-VR usage in research and application. As HMD-VR is not necessarily only a visual experience, with many VEs including auditory elements such as speech and music, it is important to be able to understand how these different sensory inputs may interact, or differently impact WML. The phonological loop and visuospatial sketchpad of the multicomponent model facilitates this understanding more than both the multi-store model, and other contemporary models of working memory. For example, Cowan's embedded-processes model [82–86] suggests that the contents of 'working memory' is the 'focus of attention' within an activated portion of long term memory (figure 1.9). The individual is only consciously aware of the current focus of attention, in which information (including those from sensory inputs) can be integrated and manipulated, whilst the remaining activated portion of long-term memory remains unconscious but available for processing. However, despite the functional similarities with the multi-component model [86,87], the embedded-processes model does not distinguish between the sensory inputs to the same degree [88], rendering it less useful in the current context.

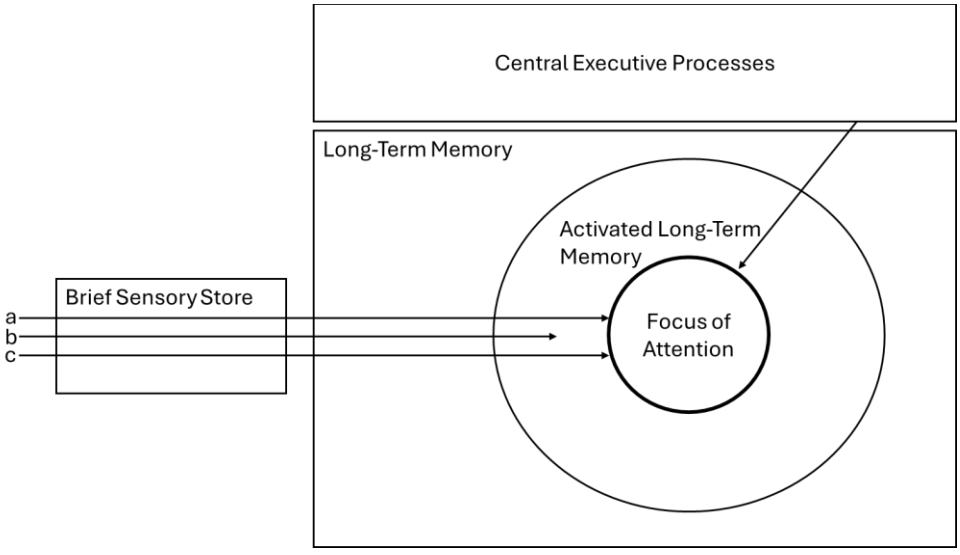


Figure 1.9: *A simplified version of the embedded-processes model of working memory. Sensory input a represents attended to stimuli in the environment. Sensory input b represents habituated, unchanging stimuli in the environment. Sensory input c represents novel stimuli in the environment. The focus of attention within long term memory is controlled by the central executive. Figure is adapted from Cowan [89].*

Working memory is limited by the amount of information that can be held online, processed and manipulated concurrently. Working memory capacity is limited to between 4-9 items or combined ‘chunks’ of information at a given time, dependant on the type of information held and individual differences between people [90,91]. It is therefore possible to overload working memory through excessive information or external factors, which can reduce performance in tasks relying on the same working memory systems [92]. Moreover, distractions and disruptions can reduce working memory performance external to the task, particularly when cognitive resources are not available to suppress the distraction [93]. It is through the reduction of external distractions, along with the control of the sensory experiences and information presented to an participant or user, that HMD-VR offers the best chance of benefitting experienced WML.

1.4.2) Cognitive Load Theory

The multicomponent model of working memory explains how different sources of information can be combined to increase experienced working memory load; however this is not a complete description of how these factors interact. Instead, Sweller’s 1988 Cognitive Load Theory (CLT) [94] builds upon Baddeley and Hitch’s concept of working memory to describe the dynamics of how different types of stimuli and information can interact to effect experienced WML. CLT was originally formulated to improve educational instruction through understanding the limitations of the working memory system, but has since developed and is now used to describe how different levels and types of experienced cognitive load can effect working memory processes [81,95–100]. CLT suggests that total experienced WML is a composite of the three subtypes of load: intrinsic load; extraneous load; and germane load.

Intrinsic load is the inherent amount of cognitive processing required to process information or complete a task, resulting from the number of items and operations that must be considered simultaneously (‘element interactivity’) [97,101,102]. For

example, in arithmetic 2+2 is relatively easier than 256+248, as the latter has more digits and required operations (such as carries between digits) that must be completed to reach the answer. Germane load refers to the amount of processing required in cognitive systems dedicated to the formation of schemas to commit to long-term memory, and is dependent on individual differences [81,97].

Extraneous load refers to task-irrelevant difficulty introduced by distractions or substandard instructional design. For example, simultaneously presenting the same information both visually and auditorily increases the amount of required processing without introducing new information [103,104]. Alternatively, attention-grabbing elements which distract from a task increases load as these must be actively supressed [105]. When working memory resources are commandeered by extraneous load, it reduces the amount available for necessary processing, hampering learning or task performance [104].

It is through the reduction of extraneous load that the benefits of HMD-VR in research and application relative to other display methods can be understood. DB-VR screens are often surrounded by distracting noises, sights and devices [106], which can be reduced by isolating individuals within an immersive VE using HMD-VR. In comparison to CAVE presentation, whilst both methods can allow participants to interact with VEs in naturalistic ways, HMD-VR minimises or prevents distracting visual disparities between real and virtual elements that can occur in projection-based VR [42].

1.4.3) Measures of Working Memory Load

As it is difficult to separately measure intrinsic, extraneous and germane loads, WML is typically measured as a composite of each type of cognitive load [107]. There are a range of behavioural, qualitative, physiological and neurological measures utilised in CLT research [108] which are often used together to produce more accurate readings of experienced load [109]. Common measures in psychological research include objective behavioural measures such as task completion time, response

accuracy and learning outcomes, with increased WML typically reducing performance on these scales [110,111]. A popular subjective measure is the self-report NASA Task Load Index (NASA-TLX) for measuring experienced cognitive workload over subscales of mental demands, physical demands, temporal demands, evaluation of own performance, effort, and frustration [112,113].

Cognitive load has also been measured and compared using objective physiological measures, including eye movement [111], eyeblinks [114], pupil dilation [115] and galvanic skin responses [114,116]. Objective neurophysiological recording measures, such as EEG and functional near-infrared spectroscopy (fNIRS), directly measure physiological responses of working memory processes occurring within the brain. For example, increased blood oxygen level dependant (BOLD) signal responses can indicate increased levels of WML, as measured using methods such as functional magnetic resonance imaging (fMRI) [117] or fNIRS [118,119]. Moreover, when using fMRI to measure BOLD signal responses, researchers can localise activation associated with working memory processes, which is typically reported in the frontoparietal regions [120].

Working memory research also utilises a range of EEG measures [108,121]. Within time-frequency analyses, changes in frontal theta and parietal alpha rhythms are commonly examined, with alpha-band desynchronisation (~8-12Hz) and theta-band synchronisation (~4-7Hz) typically representing higher levels of WML [121–127]. Changes in the synchronisation of frequency bands can be examined individually through event-related desynchronisation/synchronisation [128], or using a power ratio between the frequencies [129]. Beta activity is also targeted in CLT studies, with increased cognitive load being represented by beta band desynchronisation [122,127] in the frontal [128] and temporal regions [130]. However, it has been found that both higher and lower beta band activity can sometimes positively correlate with experienced cognitive load [122].

ERP studies using working memory tasks will either directly manipulate WML through the presented stimuli, or will measure the same stimuli when levels of WML are affected by another task [25,131]. A common target in working memory research is the ‘P300’ positive peak that occurs between 200-400ms post stimulus onset, the amplitude of which is associated with changes in working memory related processes [132]. The P300 has been associated with a range of cognitive processes including the allocation of cognitive resources to presented stimuli [133]. In the context of working memory, it is commonly found that the P300 will reduce in peak and mean amplitude when under higher levels of task-induced load [124,134]. Similar modulation under increased WML has been found with other ERP components, for example the N170 response commonly associated with facial processing has been demonstrated to reduce in amplitude when WML is increased [135,136].

1.4.4) Head-Mounted Display Virtual Reality in Studies of Working Memory

Working memory is linked to a wide range of cognitive functions [137], many of which are the subject of experiments conducted using HMD-VR. For example, HMD-VR has been used to present studies of education and learning [138–140], attention [141], memory span [142], vehicle operation [31,32], procedural knowledge [143], management [144,145], and spatial navigation [146]. HMD-VR is also being utilised in education and training contexts (see Jensen and Konradsen [147] for a review), for example to increase engagement whilst reducing distractions [148].

There have also been several comparisons between the level of WML evoked between HMD-VR and alternative presentation methods. However, a consensus on if HMD-VR increases or decreases experienced WML relative to other displays, and what factors impact this, has not yet been reached. Many of these comparisons find the use of HMD-VR benefits working memory processes, reporting the more immersive display reduces measures of WML. For example, measures of learning outcomes and skill transfer in a block-based learning task presented using non-modern DB-HMD-VR was increased compared to the DB-VR and real-life conditions [149]. Roldán et al. [145] compared performance in a management task between modern HMD-VR and DB-VR, and found performance increased when within the

HMD-VR condition. Moreover, it has been found that both learning tasks presented using smartphone HMD-VR [150] and non-modern DB-HMD-VR [151] improve learning outcomes relative to ‘real-world’ lecture-based presentations using PowerPoint. Several potential explanations have been proposed as to why HMD-VR increased working memory processes, including the higher immersion allowing for easier information processing [6], or the increased attention from excitement and engagement of using HMD-VR devices [151].

There are also several comparisons between HMD-VR and alternative display methods that find the more immersive displays increase experienced WML. Makransky et al. [16] found that, when presenting the same VE between HMD-VR and DB-VR, using HMD-VR increased neurophysiological measures of WML and task performance. Parong & Mayer [105] found that, without additional intervention to reinforce learning, the use of a modern HMD-VR decreased post-test learning outcomes compared to ‘real life’ lecture-based slideshow presentations. Frederiksen et al. [152] found that HMD-VR induced higher levels of load compared to a DB-VR simulator for laparoscopic surgery training. Each of these papers suggest that the increased WML resulted from distractions introduced by HMD-VR usage, which increased extraneous load. Both Makransky et al. [16] and Parong & Mayer [105] suggest that the use of HMD-VR itself is the source of distraction, through the increased perceptual realism or distracting levels of excitement respectively, thus increasing extraneous load. Frederiksen et al. [152] expanded upon this, suggesting that immersive VEs contain higher levels of element interactivity as participants expect to be able to interact with all the contents of a VE regardless of task relevance.

As HMD-VR is an emerging technology in psychological and neurophysiological research, the exact relationship between the display method and working memory processes is currently unknown. To date, there has been no systematic review comparing the amount of WML evoked by HMD-VR usage generally or within certain subcategories of HMD-VR relative to other presentation methods. Moreover, many high-immersion HMD-VR configurations have seen little use in working memory

paradigms, for example HS-HMD-VR models like the Vive Pro. It is important these gaps in knowledge are addressed to better understand the use of HMD-VR in psychological research, as using a presentation method which heightens WML could introduce unwanted confounding factors to experiments.

1.4.5) Virtual Reality and Cybersickness

A potential problem when utilising any form of VR in research is cybersickness [153]. Cybersickness, also called virtual reality induced symptoms and effects [154,155], VR sickness [156], or simulator sickness [157], is characterised by feelings of nausea, dizziness, disorientation, and headaches during VR usage. It is believed that cybersickness results from the 'sensory mismatch' between the visual system receiving input indicating movement, whilst the vestibular system indicates the body is remaining stationary [157–159]. Levels of cybersickness are typically measured using the simulator sickness questionnaire (SSQ), which probes total cybersickness and contains subscales of nausea, oculomotor discomfort, and disorientation [160].

Whilst the presence of cybersickness is detrimental to any study, research has suggested that cybersickness is most prevalent in VR research when using HMD devices [161]. The exact increase in cybersickness symptoms within HMD-VR experiments is inconsistent between studies (see Jensen & Konradsen [147] for a review). Some studies report <4% of participants experiencing symptoms [162], whereas others report as many as ~73% [155] or ~91% [163] of participants experiencing small-to-large increases in symptoms.

It is important to take steps minimise the negative effects of cybersickness in HMD-VR research to minimise negative symptoms experienced by participants during an experiment. Basic steps that can be taken include ensuring the HMD is properly calibrated for each participant. Stanney et al. [164] found that incorrectly calibrated interpupillary distance (IPD) of the lenses in the HMD system can result in increased cybersickness symptoms. A calibration procedure can be implemented into

experiments by including a clarity check at the start of experiments, for example having participants read a short passage presented in the VE.

Cybersickness is particularly detrimental to working memory-based studies, as it increases levels of extraneous load [165]. Both Roettl & Terlutter [141] and Oberhauser et al. [32] suggested that the increased WML observed in HMD-VR over DB-VR partially resulted from cybersickness. It is therefore important to integrate measures of cybersickness, such as the SSQ, into studies utilising HMD-VR for two reasons. Firstly, to track symptoms that participants may experience and to identify improvements that can be made in future studies to minimise symptoms. Secondly, if increases in cybersickness can introduce a confounding variable into any working memory task performed, it will be important to identify the level of sickness when interpreting the results.

1.5) Aims of this Thesis

The overall aim of this methodological thesis is to investigate the combined use of HMD-VR and EEG in the study of working memory processes. The specific aims of this thesis are explored over four studies:

1) To undertake a systematic review of the published literature to determine the utility of combined HMD-VR and neurophysiological methods in the study of working memory tasks compared to non-HMD-VR presentation methods.

2) To acquire event related potential components (P300, N170, slow wave) to a visual and auditory working memory arithmetic task using combined high-specification HMD-VR and EEG.

3) To compare the data analysis preprocessing decisions for artifact removal on ERP responses when using combined high-specification HMD-VR with EEG.

4) To compare high-specification HMD-VR with DB-VR using behavioural measures of working memory load, EEG measures of working memory load, and measures of cybersickness in a spatial navigation maze learning and recall task.

Chapter 2) A Systematic Review of the Utility of Combined Head Mounted Display Virtual Reality and Neurophysiological Recording Methods in the Study of Cognitive Workload Tasks Compared to Non-Head Mounted Display Virtual Reality Presentation Methods

2.1) Introduction

Modern HMD-VR presents the opportunity to conduct working memory research in ecologically valid VEs, allowing researchers to control the presentation of visual and auditory stimuli or scenarios whilst excluding external distractions [6,11,13,20]. However, there are several factors regarding the use of HMD-VR in the study of working memory, both alone and in combination with neurophysiological measures, which are currently underexplored or contested within the literature. These factors must be examined to fully understand the utility, including the potential advantages and disadvantages, of using HMD-VR as a research methodology.

2.1.1) Inconsistent Working Memory Load During Head-Mounted Display Virtual Reality Use

A major limitation of the current literature is that the complex relationship between HMD-VR and experienced WML is not understood. Currently, there is no consensus on if using HMD-VR reduces or increases the amount of experienced WML compared to other display methods. In comparison studies with less immersive display technologies, it has been found that using HMD-VR can increase working memory performance and reduce experienced WML compared to alternative displays methods [6,145,150,151]. Conversely, other studies report that the use of higher-immersion VR or HMD-VR can instead increase measures of WML relative to other forms of VR presentation [16,32,152,166].

Evidence from experiments comparing WML between HMD-VR and a lower-immersion display using similar paradigms have found opposing results of which evoked a lower WML. Both Ray & Deb [150] and Parong & Mayer [105] utilise learning outcome tasks to compare HMD-VR with a traditional classroom, but find contrasting results on which presentation method resulted in a lower WML. The conflicting findings leave it uncertain if certain paradigms or cognitive processes benefit from HMD-VR usage. As many of the psychological paradigms and real-world applications are based in working memory or adjacent principles [138], it is important to identify broadly if HMD-VR is detrimental to experienced load. If HMD-VR inherently increases WML relative to other display methods, it will limit the use of the technology in research and application.

2.1.2) Potential Problems When Combining Head-Mounted Display Virtual Reality and Neurophysiological Recording in Working Memory Research

2.1.2.1) The Introduction of Noise to the Recorded Neurophysiological Signals

As discussed in section 1.4.4, there are several neurophysiological measures that can be used alongside HMD-VR to provide an objective measure of experienced working memory load. Certain head-mounted neurophysiological recording methods, such as EEG [63] and fNIRS [119], can be placed underneath many consumer-grade HMDs with little-to-no modification, making them a practical choice for research. However, there are concerns regarding the utility of combined HMD-VR and neurophysiological recording techniques due to the introduction of noise to the neurophysiological signals, particularly EEG [167]. For example, Hertweck et al. [56] demonstrates how modern DB-HMD-VR can introduce line noise to an EEG signal due to the proximity of the devices, and reports that electrical noise recorded in EEG can differ between HMD-VR configurations. This electrical noise is problematic when analysing the data, reducing the quality of the recorded data and potentially rendering it unusable.

2.1.2.2) Discomfort and Cybersickness

An additional concern for combining HMD-VR and neurophysiological recording methods arises from the potential for discomfort, particularly that resulting from the

weight and pressure placed on the head when wearing the devices. Discomfort has been linked to experienced cybersickness [168], which can distract from the task and thus increase extraneous load [165]. In severe cases, it can prevent participants from completing the experiment, or otherwise negatively impact participant behaviour or recorded neurophysiological signal.

Despite cybersickness being an important factor in HMD-VR research that may impact the interpretation of the results, papers often do not examine or report experienced cybersickness. The EEG studies referenced in Chapter 1 which compared between HMD-VR and DB-VR in [16,105] found results indicating that HMD-VR had decreased working memory task performance, indicating an increased WML. However, neither paper reported anecdotal discomfort or measures of cybersickness nor suggested either contributed to the findings, potentially overlooking an explanation for their findings. Moreover, as cybersickness results from several factors relating to immersion, such as FOV and smooth navigation [159], the increased technical specifications and ergonomic improvements of modern HMD-VR may counteract the potential discomfort of combining methods.

2.1.3) Aims of this Systematic Review

The advent of modern HMD-VR presents exciting opportunities for working memory research, however there is still many questions surrounding its use as a research methodology in comparison to established display methods. There are conflicting findings regarding the use of HMD-VR on experienced WML relative to other display methods, with it currently being uncertain what, if any, impact HMD-VR has on working memory processes. Moreover, the level of immersion provided by HMD-VR differs between HMD-VR configurations, which in turn may differently effect experienced WML. It is also important to understand how HMD-VR has been used with neurophysiological measures of WML, to identify how the methodologies have been combined and what compatibility issues have been encountered. As neurophysiological recording methodologies are an important tool in psychology, the possibility of HMD-VR interfering with the capturing of data could preclude the combined use in research. Taken together, understanding these aspects of HMD-VR

in research will elucidate the potential utility of the display in wider working memory research and application.

This systematic review aims to explore the utility of HMD-VR as a tool in working memory research as both a method of presenting experimental stimuli, and specifically when combined with neurophysiological recording techniques. Papers comparing HMD-VR presentation of working memory tasks to other presentation methods, including other forms of VR and real-life equivalents, as measured by neurophysiological recording techniques, will be systematically identified and reviewed. The first objective is to identify whether using HMD-VR affects working memory processes differently compared to other presentation methods, particularly investigating if it increases experienced WML. The investigation will include identifying what HMD configurations have been used in research, and which specific factors of an HMD-VR configuration or experience have been reported or speculated to effect experienced load. The second objective is to identify what categories of HMD-VR have been successfully combined with neurophysiological recording methodologies, and identify any compatibility issues between devices which have been identified. To the researcher's knowledge, this review is believed to be the first of its kind, and will offer suggestions to guide future research.

2.2) Methods

2.2.1) Selection of Search Terms and Databases

2.2.1.1) Summon Review

The first 'Summon' search was conducted using the University of Hull's (UoH) Summon system, which probes databases across all research fields. The search targeted papers which compare HMD-VR with another display method in a working memory or cognitive load task using a neurophysiological measure of WML. Papers which were not included in the Summon databases but were available elsewhere online in external databases were included where possible. The final search terms and parameters used were the product of eight revisions.

Three key papers comparing HMD-VR presentation to an alternative display method whilst measuring a form of WML, Makransky et al. [16]; Oberhauser et al. [32]; and Webster [151], formed the basis of the search terms used. Whilst these papers do not necessarily use neurophysiological recording methods, they otherwise contained suitable comparisons and included key terms for *"Virtual Reality"* and *"Cognitive Load"*. The final search was required to return the three key papers, as it would indicate that appropriate paradigms were captured. Neurophysiological recording terminology was not used here to prevent unintentionally excluding any methodologies.

The first iteration, *(head-mounted display) AND (cognitive load)*, outlined the fundamental terms of the search, dividing the search between the method and the measure. These terms are then expanded upon with each iteration, for example revision 2 used *(head-mounted display) AND ((cognitive load) OR (workload)),* to expand for alternative ways of referring to working memory tasks.

The third revision compared using virtual reality in the search terms against filtering results by the 'subject terms' function of the search engine. Revision 3.1 introduced *AND (virtual reality)* to the end of the search to exclude non-VR papers or papers focusing on AR, whilst revision 3.2 added the field of 'Virtual Reality' to the subject filter in the summon system. However, 3.2 was found to exclude Oberhauser et al. [32] and Webster [151], thus 3.1 was selected to reduce the chance of suitable papers being erroneously excluded.

Revision 4 added *OR (head mounted display) OR (HMD)* to the VR statement. This expanded the number of returned papers from 329 to 1176, however excluded Oberhauser et al. [32] and Webster [151] despite OR statements being used. Revision 5 introduced quotation marks around each term to exclude irrelevant papers, but did not reintroduce the two excluded key papers.

Revision 6 was a series of 4 attempts to refine the search terms to ensure the key papers were captured. This included attempts at expanding both the VR terminology with *“Simulator”* (6.1) and *“Immersive”* (6.2), and the methodology terminology with *“Learning Outcomes”* (6.3) both individually and in combination. When *“Immersive”* and *“Learning Outcomes”* were used in combination (6.4), each key paper was included. Including *“Simulator”* provided no benefit and was not used in the final version.

Revision 7 provided the final search terms used in the Summon Searches. *“Immersive”* was changed to *“Immersive Virtual Reality”* upon reviewing the number of false-positive papers, reducing the number of papers from 1910 to 872. An attempt to reduce the number of reviews by including *AND ("Measurement")* to the search terms was tested, however this excluded two key papers and was ultimately unused.

The final search terms used in the Summon search are broken into three inclusive-OR sections, connected by AND statements: Firstly, common HMD naming conventions were included using *(("head-mounted display") OR ("head mounted display") OR ("HMD") OR ("Immersive virtual reality"))*; secondly, *(("cognitive load") OR ("workload") OR ("Learning Outcomes"))* were used to capture measures of cognitive load and working memory; thirdly, *(“virtual reality”)* was included to better exclude non-VR papers. The final revision used in the summon search, revision 8, updated the inclusion criteria to only include papers that had full text available online in English. The first search conducted was in May 2019 capturing 808 papers; and a follow-up search was conducted during the COVID-19 pandemic in April 2020 using the same parameters and search terms, capturing an additional 176 papers, totalling 984 results.

2.2.1.2) EBSCOhost Review

Upon review of the papers collected during the Summon search, it became apparent that many working memory tasks which did not explicitly refer to cognitive load or

workload were excluded from the search results. Thus, a refocused search was performed from relevant databases (Psychinfo, PsychArticles, Education Research Complete, and MEDLINE) using the EBSCOhost search engine. Modified search terms (revision 9) changing the “Cognitive Load” statement to *(("Cognitive") OR ("Workload") OR ("task") OR ("paradigm") OR ("memory") OR ("attention"))* was used to better include working memory tasks. Neurophysiological recording terms identified during the Summon Searches were included to limit irrelevant returns, using *(("behavioural") OR ("neuroimaging") OR ("imaging") OR ("EEG") OR ("ERP") OR ("fNIRS") OR ("fMRI"))*. Whilst the first search only captured EEG and fNIRS papers, *“fMRI”* and *“behavioural”* were included at the search engine’s recommendation. The inclusion criteria of the three initial papers were dropped as not all utilised neurophysiological recording methods. The search was performed in August 2020, with 196 papers being identified.

The EBSCOhost search was verified through comparison to a matched Summon search. When using the updated search terms and restricting results to psychology, education and medicine databases within the Summon system, it was found certain relevant papers captured in the EBSCOhost search, such as Cho et al. [169], were not returned in a search despite being present in Summon’s archives. Therefore, the EBSCOhost search was deemed appropriate for inclusion, and was combined with the two Summon searches to provide a comprehensive review of the available research, forming a total combined list of 1180 papers. The complete revision notes of both sets of search terms can be seen in Table 2.1.

Table 2.1: *Breakdown of the evolution of the search terms used in this review. New additions compared to the previous accepted version are highlighted in green. Any variation that was not used is shaded in light grey.*

Revision	Search Terms	Search Engine	Search Engine Criteria	Number of Results	Excluded Key Papers	Date
1: Basic search	(head-mounted display) AND (cognitive load)	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	1256	None	May 2019
2: Adding workload	(head-mounted display) AND ((cognitive load) OR (workload))	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	1781	None	May 2019
3.1: Specifying VR	(head-mounted display) AND ((cognitive load) OR (workload)) AND (virtual reality)	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	898	None	May 2019
3.2: Specifying VR (Rejected)	(head-mounted display) AND ((cognitive load) OR (workload))	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed 3) Subject Terms: Virtual Reality	329	Oberhauser et al. [32] Webster [151]	May 2019
4: Expanding HMD	((head-mounted display) OR (head mounted display) OR (HMD)) AND ((cognitive load)	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	1176	Oberhauser et al. [32] Webster [151]	May 2019

	OR (workload)) AND (virtual reality)					
5: Tightening the search	(("head-mounted display") OR ("head mounted display") OR ("HMD")) AND (("cognitive load") OR ("workload")) AND ("virtual reality")	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	518	Oberhauser et al. [32] Webster [151]	May 2019
6.1: Including the key papers: Simulator (Rejected)	(("head-mounted display") OR ("head mounted display") OR ("HMD") OR ("Simulator")) AND (("cognitive load") OR ("workload")) AND ("virtual reality")	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	1647	Webster [151]	May 2019
6.2: Including the key papers: Learning	(("head-mounted display") OR ("head mounted display") OR	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	646	Oberhauser et al. [32] Webster [151]	May 2019

Outcomes (Rejected)	("HMD")) AND (("cognitive load") OR ("workload") OR ("Learning Outcomes")) AND ("virtual reality")					
6.3: Including the key papers: Immersive (Rejected)	(("head-mounted display") OR ("head mounted display") OR ("HMD") OR ("Immersive")) AND (("cognitive load") OR ("workload")) AND ("virtual reality")	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	1291	Webster [151]	May 2019
6.4: Including the key papers: Immersive and Learning Outcomes	(("head-mounted display") OR ("head mounted display") OR ("HMD") OR ("Immersive")) AND (("cognitive load") OR ("workload") OR ("Learning	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	1910	None	May 2019

	Outcomes")) AND ("virtual reality")					
7.1: Tightening the search again: Immersive VR	(("head-mounted display") OR ("head mounted display") OR ("HMD") OR ("Immersive virtual reality")) AND (("cognitive load") OR ("workload") OR ("Learning Outcomes")) AND ("virtual reality")	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	872	None	May 2019
7.2: Tightening the search again: Immersive VR and Measurement (Rejected)	(("head-mounted display") OR ("head mounted display") OR ("HMD") OR ("Immersive virtual reality")) AND (("cognitive load") OR ("workload") OR ("Learning Outcomes"))	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed	431	Oberhauser et al. [32] Webster [151]	May 2019

	AND ("Measurement") AND ("virtual reality")					
8: Full text available online	(("head-mounted display") OR ("head mounted display") OR ("HMD") OR ("Immersive virtual reality")) AND (("cognitive load") OR ("workload") OR ("Learning Outcomes")) AND ("virtual reality")	Summon (UoH)	1) Outside Library Results Included 2) Peer Reviewed 3) Full Text Online	799	None	May 2019
9: EBSCOhost Search	(("head-mounted display") OR ("headset") OR ("headsets") OR ("head mounted display") OR ("immersive virtual reality") OR ("HMD")) AND ("virtual reality") AND	EBSCOhost (APA PsychArticles, APA PsycInfo, MEDLINE, Education Research Complete)	None	196	N/A	Aug 2020

	(("Cognitive") OR ("Workload") OR ("task") OR ("paradigm") OR ("memory") OR ("attention")) AND (("behavioural ") OR ("neuroimaging ") OR ("imaging ") OR ("eeg ") OR ("erp ") OR ("fnirs ") OR ("fMRI"))					

2.2.2) Inclusion Criteria

The papers captured were accepted or rejected based on three criteria. Firstly, the paper must compare HMD-VR to an alternative presentation method by evoking working memory in a comparable way. For example, acceptable learning outcome paradigms must present the same information between display conditions. The alternative display can either be another form of VR such as SB-VR or CAVE, a form of augmented reality, or a 'real life' alternative such as a lecture setting or physical task. In instances where there are several comparisons, such as presenting a task completed in a real-life environment to the same task in both an HMD-VR simulation of the same laboratory, or a different environment entirely (e.g. Peterson et al. [25]), only the most balanced comparison will be considered. Secondly, there must be a neurophysiological measure of WML, process associated with working memory, or process associated with cognitive load. By extension, behavioural results considered must directly or indirectly measure working memory or cognitive load using common methodologies such as task performance, learning outcomes, or subjective scales.

Thirdly, non-experiment papers such as reviews or book chapters not including an experiment, non-English papers, duplicate results, or papers that are otherwise inaccessible are excluded.

2.2.3) The Review Process

Each paper returned by the search terms above was scrutinised based on the inclusion criteria to identify suitability for this review over a three-stage process. The review process and subsequent analysis is based on contemporary systematic reviews which examined comparisons of working memory with a focus on learning or virtual reality technologies [170–172].

2.2.3.1) Step 1: Initial Abstract Scan

The purpose of the first step of the review process was to identify which research papers used a form of HMD-VR in a working memory or cognitive load task, and remove any that do not. Each accessible paper was downloaded in a PDF format, and the title and abstract were examined to identify the task and methodology used. The body of the text was searched for key words indicating HMD usage (‘HMD’, ‘Immersive’, etc.), and the methods section was reviewed for the apparatus and task utilised. Papers that obviously did not use HMD-VR, had no indication of a comparison of working memory, or otherwise did not meet the third set of inclusion criteria were excluded. Papers that explicitly used these methods, or were ambiguous on the methodology, were permitted to the next step to avoid unnecessarily rejecting potentially suitable studies.

2.2.3.2) Step 2: Full Read and Data Extraction

Each paper that progressed to the second stage underwent a full read and was subjected to the full inclusion criteria to ensure suitability for this review. In addition to the criteria used in Step 1, any paper that did not perform comparisons between HMD-VR and an alternative display method, did not use a working memory or cognitive load task, or did not include an appropriate neurophysiological measure of working memory or cognitive load was rejected. Accepted papers had relevant key

factors extracted: the headset and input methods used; the modality the HMD was compared against; sample sizes; the task; conditions; measurements; findings; conclusions and relevant contextual notes.

2.2.3.3) Step 3: Snowballing and List Combination

For each paper accepted during Step 2, two additional ‘snowballing’ searches were performed to identify papers with relevant titles for this review from outside the search databases. Backwards snowballing was conducted by examining the reference list of accepted papers, whilst forwards snowballing was performed using the ‘cited by’ function of Google Scholar. Forwards snowballing was also performed on relevant papers that were rejected in Step 2 for not including a comparison between display methods, but did include neurophysiological measures of cognitive load. Titles were considered relevant if they included a variation of ‘virtual reality’ and a working memory associated cognitive process. Additional papers identified by the snowballing processes underwent the process as outlined in Step 2, and accepted papers were included in the final list. Additionally, any inter-search duplicates between the Summon and EBSCOhost searches were identified and removed.

2.3) Results and Discussion

The present review systematically collected and examined neurophysiological studies comparing task-induced WML between HMD-VR and alternative display methods. The aim of this systematic review was to explore the utility and suitability of combined HMD-VR and neurophysiological recording methods in working memory research, as divided between two objectives. The first objective was to identify if using HMD-VR introduces higher levels of WML over comparison presentation methods, as measured using neurophysiological recording methods. The second objective was to identify what categories of HMD-VR have been combined with neurophysiological recording methods, and what the shortcomings of combining the technologies may be in the context of the neurophysiological recording, experienced WML, and participant comfort.

2.3.1) Included Papers

The combined Summon searches identified 984 papers and the EBSCOhost search identified 196 papers, producing an initial list of 1180 papers. In the first review step, 151 Summon papers and 38 EBSCOhost papers progressed, totalling 189. Of the 991 rejected papers: 608 research papers did not use HMD-VR or contain an appropriate comparison; 285 were reviews or otherwise not original research publications; 65 papers were duplicates within the search (23 in the Summon searches and 42 in the EBSCOhost search); 40 papers only had the abstract available; and 5 were not available in English.

The second step accepted 7 Summon papers and 9 EBSCOhost papers, totalling 16. Of the rejected 173 papers, 166 papers were excluded for not containing an appropriate comparison of display methods, a working memory task, or a neurophysiological recording method. A total of 6 non-experiment papers and 1 duplicate of a previously rejected paper with a different citation were also removed.

An additional 15 titles were identified through snowballing. Only 1 paper was accepted for inclusion, as the remaining 14 papers did not employ an appropriate display method comparison, cognitive task or neurophysiological recording method. All accepted papers were compiled into the combined list of 17 and reviewed for any duplicate entries. Two pairs of duplicates were identified between searches and one copy of each was removed. The complete review process can be seen in Figure 2.1. In total, 15 papers were included in this review, with the full list detailed in Table 2.2.

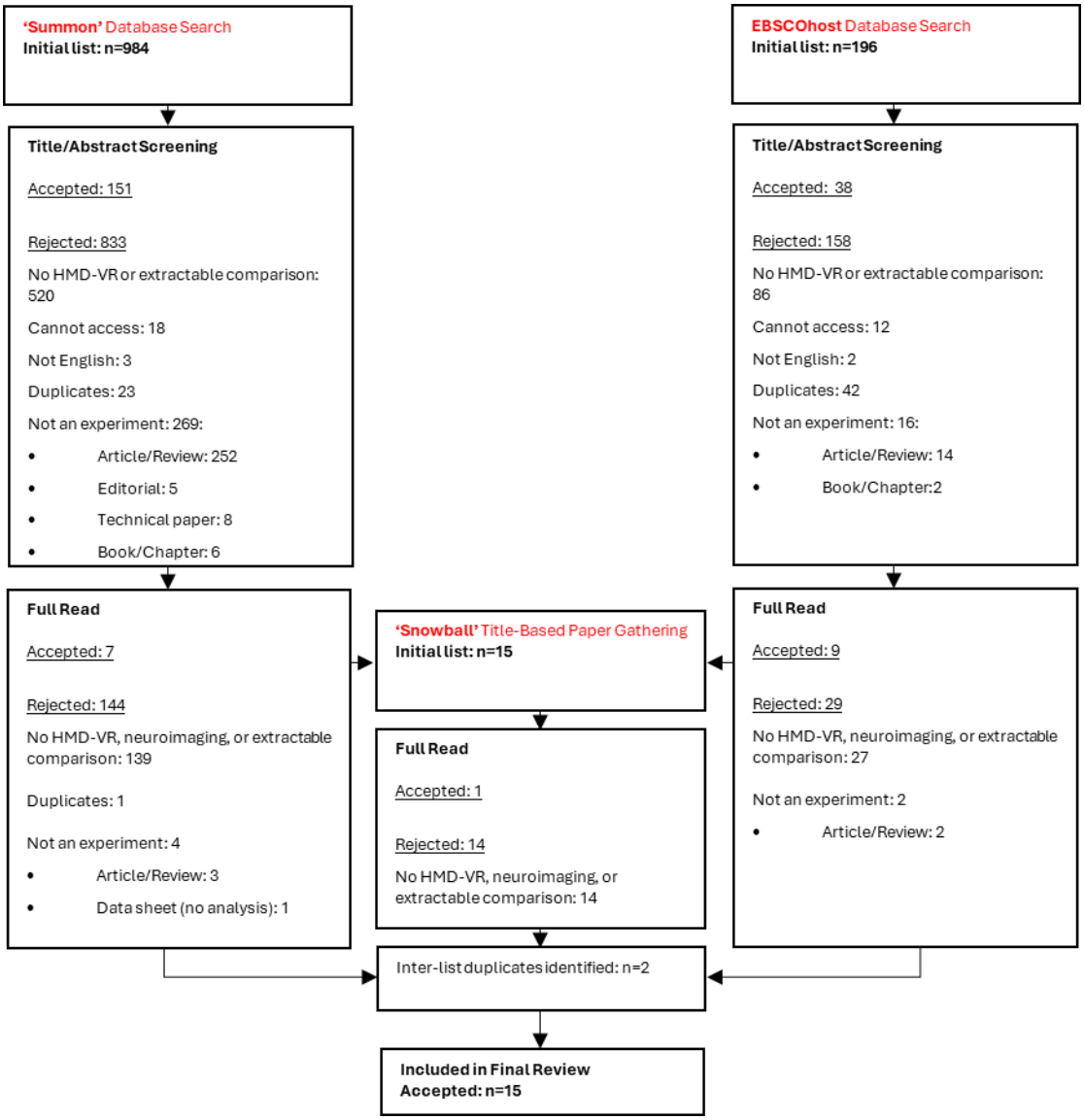


Figure 2.1: *A flow diagram of the complete review process of this systematic review, including the number of accepted papers and reason for rejected papers for both searches.*

Table 2.2: *Information about each paper accepted by this systematic review, including which HMD-VR device was used, what the comparison display is, what type of working memory task was used, what type of neurophysiological measure was used, and what results were reported.*

Title and Authors	Headset Used	Comparison Display	Task and Conditions	Neurophysiological Measure and Results	Behaviour Measures and Results
Use of auditory event-related potentials to measure immersion during a computer game **Burns & Fairclough [173]**	Non-Modern DB-HMD-VR: Silicon Micro Display ST1080-10V1 head-mounted display (2012)	SB-VR (Television)	Primary: Video game to manipulate WML through difficulty Secondary: Oddball task (vigilance)	EEG: ERP Mean Amplitude P1, LN, Slow Wave Deflection	N/A
				Not significantly different	N/A
Neurofeedback Training with Virtual Reality for Inattention and Impulsiveness **Cho et al. [169]**	Non-Modern DB-HMD-VR: Daeyang E&C (unidentified, likely the Cy-Visor DH-4400VP, 2000/2001)	SB-VR (Computer Monitor)	Continuous performance task	EEG: Power Spectra Cz Mean Beta Ratio	Before and after intervention/ training sessions Performance Response time "Perceptual sensitivity" (Time to completion) Omission and commission errors
				HMD-VR induced a higher Beta-Wave Ratio	Higher task performance in HMD-VR

Assess BA10 activity in slide-based and immersive virtual reality prospective memory task using functional near-infrared spectroscopy (fNIRS) **Dong et al. [174]**	Development-Grade Modern DB-HMD-VR: Oculus Rift DK2 (2014)	SB-VR (Computer Monitor/ Slides)	Prospective memory task	fNIRS: BOLD Response Left BA10	N/A
				HMD-VR had Higher activation in prospective memory areas	N/A
Manipulating bodily presence affects cross-modal spatial attention: A virtual-reality-based ERP study **Harjunen et al. [167]**	Development-Grade Modern DB-HMD-VR: Oculus Rift DK2 (2014)	SB-VR (Computer Monitor)	Attention: Comparing an unaltered view with an HMD-VR view (with hands visible)	EEG: ERP Peak Amplitudes Midline P200, N200, P3	Oddball task (Primary)
				Stronger P200 in the HMD-VR condition, all other comparisons no difference	N/A
Embodiment is related to better performance on a brain–computer interface in immersive virtual reality: A pilot study **Juliano et al. [175]**	Consumer-Grade Modern DB-HMD-VR: Oculus Rift CV1 (2016)	SB-VR (Computer Monitor)	BCI training task: imagined motor activity	EEG: Power Spectral Density Central Alpha, Beta and 8-24Hz	Neurofeedback performance Completion time Successful trials
				Not significantly different	Not significantly different

Rapid P300 brain-computer interface communication with a head-mounted display **Käthner et al. [176]**	Development-Grade Modern DB-HMD-VR: Oculus Rift DK1 (2013)	SB-VR (Computer Monitor)	BCI: P300 speller task Comparison between whole screen VR and SB-VR	EEG: ERP Mean Amplitude VPP, LPP, N170, P300	BCI Accuracy
				Not significantly different	Not significantly different
Enhanced attention using head-mounted virtual reality **Li et al. [177]**	Consumer-Grade Modern DB-HMD-VR: HTC Vive (2016)	SB-VR (Computer Monitor)	Selective attention task	EEG: Power Spectra Beta/Theta Ratio (BTR) ERSP: FZ Theta ERP Peak Latency P3a and P3b	Response Time Response Accuracy
				HMD-VR had higher markers of attention Higher BTR Higher Theta Shorter ERP latency	Higher task performance in HMD-VR
Adding immersive virtual reality to a science lab simulation causes more presence but less learning **Makransky et al. [16]**	Smartphone-HMD-VR: Samsung GearVR with Samsung S6 (2015)	SB-VR (Computer Monitor)	Learning task	EEG: Power Spectra Frontal-Central and Central-Parietal 1-40Hz	Learning Outcomes
				Higher Cognitive Load in HMD-VR based on a workload classifier	Lower learning outcomes in HMD-VR

Effects of virtual reality high heights exposure during beam-walking on physiological stress and cognitive loading **Peterson, et al. [25]**	Development-Grade Modern DB-HMD-VR: Oculus Rift DK2 (2014)	Real Life Task	Visuomotor task Comparing non-perturbed HMD-VR to unaltered (real life) presentation Secondary task: tone vigilance task	EEG: ERP Peak Latency within 500-600ms Focused on the Anterior Cingulate Cluster	Balance and number of crosses along the beam Response time to an auditory cue
				Reduced amplitude and longer latency in the HMD-VR condition	Lower task performance in HMD-VR
Transient visual perturbations boost short-term balance learning in virtual reality by modulating electrocortical activity **Peterson, et al. [178]**	Developer-Grade Modern DB-HMD-VR: Oculus Rift DK2 (2014)	Real Life Task	Visuomotor Task: Comparing Non-perturbed HMD-VR to Unaltered (real life) presentation	EEG: Power Spectra (Event Related Spectral Perturbation) Frontal Alpha, Beta, Gamma and Theta	Task performance How many times participants stepped off the beam Pre-test and post-test performance
				Found cortical processing increased in alpha in HMD-VR condition No change in theta, beta or gamma	Lower Task Performance in HMD-VR
EEG Acquisition During the VR Administration of Resting State, Attention, and Image	Consumer-Grade Modern DB-HMD-VR: HTC Vive (2016)	SB-VR (Computer Monitor)	Vigilance task	EEG: ERP Amplitude All Channels LPP Power Spectral Analysis All Channels Alpha	Task performance Response Time Response accuracy

Recognition Tasks: A Feasibility Study **Rupp et al. [66]**				Not significantly different	Not significantly different
Embodied VR environment facilitates motor imagery brain–computer interface training **Škola & Liarokapis [179]**	Consumer-Grade Modern DB-HMD-VR: Oculus Rift CV1 (2016)	SB-VR (Computer Monitor)	BCI: imagined motor activity	EEG: Power Spectra-Based Classifier Frontal-Central Mu and Beta Event Related Desynchronisation: 8-30Hz	N/A
				HM-VR offered significantly higher BCI classifier success No difference in ERD	N/A
The effect of a virtual reality learning environment on learners' spatial ability **Sun et al. [180]**	Smartphone HMD-VR: Mi6 phone with Google Cardboard (2017)	Real Life Lecture	Primary: learning task Secondary: oddball task	EEG: ERP Peak Amplitude N1 and P1	Learning outcomes (Pre/Post-Test)
				Overall: N1 and P2 was higher in the real life lecture condition, suggesting a higher load High-Spatial Ability Participants: Not Significantly Different Low-Spatial Ability Participants: Lower Cognitive Load in HMD-VR	Overall: No difference High-spatial ability: Lower learning outcomes in HMD-VR Low-spatial ability: Higher learning outcomes in HMD-VR

				(better)	
Studying the Effect of Display Type and Viewing Perspective on User Experience in Virtual Reality Exergames **Xu et al. [181]**	Consumer-Grade Modern DB-HMD-VR: Oculus Rift CV1 (2016)	SB-VR (Television)	Visuomotor gesture task	EEG: Power Spectra "Engagement Index" (Including Alpha, beta and theta) AF8/AF9 and TP9/TP10	Task Performance Gesture accuracy Exertion NASA TLX
				Not significantly different	Not significantly different
Examining creativity through a virtual reality support system **Yang et al. [182]**	Consumer-Grade Modern DB-HMD-VR: HTC Vive (2016)	Real Life - Task	Creativity task	EEG: Power Spectra Prefrontal/ Frontal 'Meditation/ Relaxation' algorithm, Attention algorithm	Creativity, as related by a series of experts and scales
				Attention: Not Significantly Different Meditation/ Relaxation: Brainwave (EEG device) algorithm indicated that HMD-VR had significantly higher focus (Better)	HMD-VR Creations were rated as "more creative"

2.3.1.1) Quality Assessment and Risk of Bias in the Included Papers

Five potential sources of bias were evaluated for each included paper. The first was category size bias from the sample and group sizes of included participants, which found that only 3 papers had 20 or more participants per group. Of the remaining papers, 3 papers and 1 study in a paper containing 2 studies [181] used between 15

to 19 participants per group, whilst the remaining had below 15 participants per group. It is therefore suggested that there is a high risk of category size bias which may limit the power of the results found, as the standard included number of participants is 20 in EEG studies (see Larson and Carbine [183]). Moreover, the recommended power calculation for number of participants required in an fNIRS study [184] is not conducted in Dong et al. [174].

The inclusion criteria of the accepted papers indicate the existence of selection bias in 3 papers where only males participants were recruited [169,174,181]. Whilst participants aged above 30 were sometimes reported [16,66,176], the average age of participants when reported was also under 30 in all included papers, likely due to the primary source of recruitment being from university students. There is a low chance of attrition bias however, as most papers did not report any participants exiting the study due to cybersickness. Those that did report early exits only had 2 in total per study [25,178].

Regarding the potential procedural bias, only one paper included did not randomise or counterbalance trials when multiple conditions were collected per participant [174], where the PowerPoint-slide condition was presented before the HMD-VR condition in all cases. Moreover, Peterson et al. [25] randomised the order of the HMD-VR condition, but always separated them with a non-VR condition. Overall, there is a low chance of order effects or similar biases present in this review. All papers also fully reported the results and detailed the neuroimaging findings in their discussions. Whilst several papers did not delve into deep discussions of the neuroimaging findings, these papers either did not place the primary focus on the neuroimaging results, or found limited significant results.

Considered together, there is a mixed indications of bias within the included papers of this review. Whilst most sources of bias are low, the low sample sizes and presence of selection bias in several papers warrants caution when attempting to generalise the results found.

2.3.2) Comparisons of Working Memory Load Between Head-Mounted Display Virtual Reality and Other Presentation Methods

To address this review's first objective, the outcomes of papers comparing neurophysiological recording measures of experienced WML or WML performance between tasks presented using HMD-VR and an alternative display method were examined. For each accepted paper, the working memory task, type of HMD-VR device, alternative display method, measures of experienced WML or WML performance used, and the overall findings reported were extracted. The extracted features are then considered together to understand how using HMD-VR influenced experienced load both against non-HMD-VR collectively, and against specific display methods. The interpretation of the findings is based on the authors' conclusions of the results, and discussion of the associated cognitive functions of the reported neurophysiological response.

2.3.2.1) Neurophysiological Results Between Head-Mounted Display Virtual Reality and Non-Head Mounted Display Presentations

To understand how HMD-VR compares to other display methods generally, the overall neurophysiological results for the comparison of cognitive load between display conditions is considered. Of the 15 accepted papers, 14 utilised EEG and 1 used prefrontal fNIRS to measure experienced WML. Of the 15 papers, 7 papers (6 EEG and 1 fNIRS) found that overall HMD-VR conditions had significantly lower measures of experienced WML. Lower experienced WML in HMD-VR was found over a range of cognitive load tasks, including: a higher mean beta ratio during a continuous performance task [169]; an increased BOLD response in the prospective memory areas during a prospective memory task [174]; an increased P200 visually evoked potential during an attention task [167]; increased beta/theta ratio, theta ERSP response and a faster P3a/P3b ERP peak latency during selective attention task [177]; increased classifier success during a imagined motor activity brain computer interface (BCI) task [179]; increased N1 and P1 peak amplitude during a learning task [180]; and increased attention as determined by an algorithm during a task exploring creativity in artistic creation [182].

Five EEG papers reported no significant differences between conditions. The papers that found no differences in experienced WML included P1 and LN mean amplitude ERP during a secondary oddball task [173]; power spectral density during an imagined motor activity BCI task [175]; vertex positive potential (VPP), late positive potential (LPP), N170 and P300 mean ERP amplitude during a P300 speller task [176]; LPP ERP amplitude and alpha power spectral analysis during a vigilance task [66]; and using a power spectra engagement index during a visuomotor gesture task [181].

A total of 12 papers found HMD-VR induced comparable or decreased levels of WML relative to other presentation methods, indicating that HMD-VR generally does not increase experienced WML relative to other display methods commonly used in research. However, three EEG papers reported that participants in the HMD-VR condition experienced significantly higher WML compared to the comparison display. One paper reporting increased WML in HMD-VR used a learning task which found experienced WML was higher in the HMD-VR condition based on a 1-40Hz classifier during a learning task [16]. The remaining two papers contained visuomotor balance-beam walking tasks with working memory recordings, one reporting decreased N600 ERP peak amplitude to a tone presented in a secondary task, and the second reporting decreased frontal alpha during walking [25,178].

A tentative conclusion that can be drawn from the above results is that HMD-VR does not inherently increase experienced WML. The majority of the results examined find that HMD-VR either decreased or had no effect on experienced WML, suggesting that it may be an appropriate tool for working memory research as it does necessarily not introduce extraneous load during usage. However, the existence of papers reporting WML increased during HMD-VR usage indicate that a deeper examination is required to understand the dynamics of HMD-VR usage and experienced WML, and by extension the utility of the combined technologies in research.

2.3.2.1.1) Multiple Neurophysiological Measures of Working Memory Load in Singular Studies

Many of the papers identified in this review utilised multiple neurophysiological measures of experienced WML across EEG analysis techniques, target components and frequencies, or comparison locations. When these separate analyses are taken independently of each other, they can provide differing findings. For example, Yang et al. [182] found that whilst the neurophysiological measures of meditation/relaxation indicated HMD-VR increased focus and thus lowered experienced WML, measures of attention also representing experienced WML did not differ between displays. Nevertheless, the results do not directly suggest HMD-VR both increases and decreases load simultaneously, and there is no evidence to suggest HMD-VR negatively impacted experienced load, thus the overall trend is that WML decreased in the HMD-VR condition. Similarly, Škola & Liarokapis [179] found that the success rate of a BCI classifier increased following training in the HMD-VR condition, demonstrating the neurological benefit of HMD-VR in their learning task. However, ERD responses in the beta and mu frequency ranges did not significantly differ between the HMD-VR and SB-VR conditions. The lack of significant ERD findings does not detract from the significant results of the classifier, and ultimately HMD-VR was demonstrated to improve working memory functions.

There are also instances where different groups of participants within a study were contrasted. Sun et al. [180] compared low-spatial ability and high-spatial ability learners during a mental rotation task, defined by their ability to maintain and manipulate shapes online in working memory. Whilst the overall results suggest that HMD-VR did decrease the amount of WML experienced, only the low-spatial ability group benefitted from the additional visual aid, whereas the high-spatial ability group found no difference between displays. It is argued here that HMD-VR aiding one group but not impairing the cognitive processing of another is sufficient in indicating that HMD-VR can be used in working memory research, whilst also highlighting the impact that individual differences can have on the WML during HMD-VR usage. Researchers using HMD-VR should therefore be aware of these factors when designing experiments and selecting population samples for participation.

Harjunen et al. [167] also found HMD-VR significantly lowered experienced WML compared to SB-VR, however this was only found when comparing visually evoked P200s and was absent in the majority of components examined. The number of no reported differences between the components and comparisons, which included different times of recording (e.g. early and late training), allows an argument that this paper finds no real difference between conditions.

The least charitable interpretation of Yang et al. [182], Sun et al. [180], Škola & Liarokapis [179], and Harjunen et al. [167] would be that there is not enough evidence to suggest HMD-VR reduced experienced WML over the comparison display, and thus these papers do not demonstrate any difference. However, this interpretation of the papers does not change the outcome of this review, as none of the outcomes of these papers suggest that HMD-VR increases experienced WML compared to the comparison display. Moreover, the same criticism of insufficient evidence suggesting an increase or decrease in experienced WML can be levelled towards papers reporting increased experienced WML in the HMD-VR condition. Both Peterson et al. [25] and Peterson et al. [178] included a range of additional comparisons that found no difference between conditions. Similar to Škola & Liarokapis [179], the power spectral analysis conducted by Peterson et al. [178] only found differences in alpha-wave activity, but reported no differences in the theta, beta or gamma ranges between the HMD-VR without perturbations and unaltered real-life conditions. Furthermore, the ERP analysis conducted in Peterson et al. [178] only found significant differences in the anterior cingulate cluster, but not in any of the remaining 7 clusters including over the frontal, parietal or occipital areas recorded, similar to Harjunen et al. [167].

When the strict interpretations are applied to the identified papers, only a singular paper finds a definitive neurophysiological marker of increased experienced WML when using HMD-VR. With 14 of the 15 papers finding HMD-VR either decreased or had no effect on experienced WML, it is suggested that HMD-VR does not negatively impact working memory processes in most working memory tasks and paradigms. The purpose of this section is not to call the preliminary conclusion drawn into

question, but to demonstrate how a stricter interpretation of the results does not change the relationship between HMD-VR and experienced WML. The change in perspective of these papers does not change the outcome reached by the initial analysis, and the majority of evidence does not suggest HMD-VR inherently increased experienced WML compared to alternative displays.

2.3.2.2) Working Memory Load Tasks in Head-Mounted Display Virtual Reality Compared to Screen-Based Virtual Reality

Building upon general HMD-VR to non-HMD-VR comparisons, it is important to identify which display methods HMD-VR have been contrasted against and which working memory tasks have been used. Of the 15 papers captured in this review, 11 papers compared HMD-VR to forms of SB-VR. Of the 11 papers comparing HMD-VR to a form of SB-VR, 5 papers concluded the HMD-VR condition evoked a lower WML, 5 reported no significant differences between display methods, and 1 reported HMD-VR increased experienced WML. Of the 5 papers that reported the HMD-VR condition evoked a lower experienced WML, the tasks used included continuous performance [169], prospective memory [174], attention [167], selective attention [177] and a BCI task looking at imagined motor activity [179]. The papers that reported no difference utilised a secondary oddball task during a video game [173], two BCI tasks (imagined motor activity [175] and a P300 speller task [176]), a vigilance task [66], and a visuomotor gesture task [181]. The paper that found HMD-VR to be detrimental to experienced load was a learning outcomes task [16].

The vast majority of papers find HMD-VR does not increase experienced WML compared to SB-VR, and either had no effect or found decreased experienced WML relative to SB-VR, supporting the conclusion that HMD-VR is a viable presentation method for working memory research. Moreover, experienced WML was found not to increase in HMD-VR across a range of working memory tasks, demonstrating the breadth of working memory processes and paradigms which have been successfully studied using combined HMD-VR and neurophysiological methodologies. The findings also suggest that evidence from studies using either display method can be considered and compared together, as there is not an inherent increase in

extraneous load introduced through HMD-VR. Therefore, the demonstration that HMD-VR is not detrimental to WML processes paves the way for its expanded usage in wider working memory research and application.

2.3.2.3) Working Memory Load Tasks in Head-Mounted Display Virtual Reality Compared to Real Life Presentation

Of the 15 papers captured in this review, 4 compared HMD-VR to non-VR over 2 working memory and 2 visuomotor tasks employing working memory measures. The first working memory task, conducted by Sun et al. [180], comprised of a learning outcome comparison between HMD-VR and a PowerPoint lecture-style presentation. Neural response to a simultaneous unattended oddball tone was used to measure experienced WML. It was found that overall experienced WML decreased in the HMD-VR condition based on the EEG results, with the low-spatial ability subgroup in particular benefitting from HMD-VR presentation. The high-spatial ability subgroup did not benefit from HMD-VR however, with no differences in experienced WML detected by the neurophysiological methods. Moreover, the high-spatial ability behavioural results suggested that experienced WML was increased in HMD-VR relative to SB-VR. When the results are considered together, they are weighted towards HMD-VR lowering experienced WML relative to real-life learning, but the results do introduce the additional consideration of individual differences between participants when designing a working memory study in HMD-VR.

Yang et al. [182] compared measures of creativity along with experienced WML using a design task. Participants were provided with standard drawing equipment in the real-life condition, or a 3D virtual mannequin that could be freely decorated in the HMD-VR condition. It was found that HMD-VR improved feelings of relaxation (i.e. lower load), but found no differences between conditions for measures of attention. The behavioural measures were also increased in the HMD-VR condition, with the work being rated as more ‘creative’ by experts, suggesting overall HMD-VR usage induced a lower experienced WML.

The remaining two papers both employed visuomotor tasks where participants walked across balance beams in a real life laboratory setting compared to a VE recreation of the laboratory in HMD-VR [25], or a live-feed of the laboratory through an HMD-mounted webcam [178]. Measures of experienced WML, including responses to a secondary tone, were included during the walking task. It was found that HMD-VR was detrimental to ERP peak amplitude and time-frequency measures of experienced WML, and reduced behavioural performance relative to the real-life condition.

2.3.2.4) Comparison of Behavioural and Neurophysiological Results

Whilst this review focuses on neurophysiological methodologies, contrasting the neurophysiological results with behavioural data may highlight differences between displays not apparent using neurophysiological measures alone. Papers that recorded behavioural results typically used multiple measures, including objective measures of WML through learning outcomes [16,180] and task performance [25,169,175–178,181], or subjective measures such as the NASA-TLX [181] and 'creativity' as judged by experts [182].

Of the 15 papers collected, 11 papers reported behavioural results. Of these 11 papers, 4 reported behavioural results suggesting significantly decreased experienced WML in HMD-VR conditions, 3 reported significant increase in experienced WML, and 5 reported no significant difference. Of the 11 papers reporting behavioural results, 10 of the papers reported behavioural results that corroborated the neurophysiological findings, and did not find behavioural results that partially deviated from the conclusion drawn by the paper.

Incongruencies between behavioural and neurophysiological findings was reported by Sun et al. [180], who found a difference between the findings of the low-spatial ability and high-spatial ability subgroups. The behavioural and neurophysiological results of the low-spatial ability learners improved in the HMD-VR condition indicating a lower amount of WML processing, but the high-spatial ability learners

had significantly lower learning outcomes suggesting a higher WML processing but no difference between neurophysiological results in the HMD-VR condition. Taken together, the neurophysiological recordings suggest that experienced WML decreased in HMD-VR, but there is no general change in behavioural findings. Overall, the discrepancy between subgroups does not change the conclusion reached, but does highlight the importance of considering the effect of individual differences between participants when interpreting the results.

2.3.2.5) Head-Mounted Display Virtual Reality Does Not Directly Increase Working Memory Load

The results to this point have suggested that HMD-VR does not inherently increase experienced WML in the majority of working memory tasks and processes. However, 3 papers find that WML was higher when using HMD-VR presentation methods, indicating that there are either instances where HMD-VR increased experienced WML relative to another display, or a secondary or confounding factor independent of the presentation itself increased experienced WML. It is therefore important to examine the instances where using HMD-VR is found to increase experienced WML or inhibit learning, either to identify tasks HMD-VR is not suitable for, or technical considerations that could benefit future HMD-VR paradigms.

Makransky et al's [16] comparison of declarative learning outcomes between matched environments contrasts the other comparisons between HMD-VR and SB-VR, finding HMD-VR evoked significantly higher levels of experienced WML across both neurological measures and reduced learning outcomes. A potential explanation for this is the use of low-immersion smartphone-HMD-VR, which may have made any text difficulty to read due to a low screen resolution [32]. However, Sun et al. [180] also used a similar smartphone-HMD-VR condition in a learning task and found HMD-VR outperformed a real-life equivalent learning task. Makransky et al. [16] speculate that a potential explanation for the increased experienced WML in the HMD-VR condition is the input method used in this experiment, which consisted of a headset-directed cursor and a tactile button on the side of the headset simulating a mouse click. The head-mounted button is both novel and unintuitive, which when

compared against the more commonly used keyboard and mouse configurations in SB-VR, likely required additional working memory resources to operate. Specifically, the input methods used in the HMD-VR condition do not represent the real world actions, adding an additional step between the intention and action that requires additional cognitive resources to resolve. The working memory requirement is exacerbated as the participant was also actively learning how to use the novel input, introducing extraneous load. The paper does however serve as an important reminder to not consider 'HMD-VR' as a singular entity, but to consider how each component in the HMD-VR configuration contributes to total WML experienced.

Of the comparisons to real-life presentation, 2 of the 4 found HMD-VR increased measures of WML. The relatively even split between comparisons of WML measures during HMD-VR to real-life presentation suggests that the type of task should be the main consideration during HMD-VR experimental design. Cognitive tasks, such as those conducted by Sun et al. [180] and Yang et al. [182], benefitted from the use of HMD-VR, whilst visuomotor tasks had increased WML and reduced performance [25,178]. The findings do not directly contest the visuomotor task comparing HMD-VR to SB-VR [181], which reported no differences between display conditions. However, it is suggested by Peterson et al. [178] that the results may in part stem from the technical limitations of the DK2 and supporting hardware used. Despite being classified as a modern DB-HMD-VR device, the DK2 is on the lower end of technical specifications. In particular, the relatively low 100 degrees FOV and low 30Hz refresh rate of the webcam utilised, which is only a third of the 90Hz recommended by Oculus VR LLC [185] and thus risks increasing cybersickness, is identified as a potential reason for the reduced performance. Wide peripheral vision and smooth visual flow (i.e. no noticeable delay between frames) is important for maintaining display stability and reducing cybersickness [186,187]. The difference between immersion can be seen when contrasted against Xu et al. [181], who successfully utilised the higher-specification Oculus Rift CV1 DB-HMD-VR in a visuomotor task without the same negative results.

2.3.2.6) Is Head-Mounted Display Virtual Reality Suitable for Research Compared to Non-Head-Mounted Display Presentation Methods?

Taken together, the results indicate that HMD-VR does not inherently introduce a higher level of experienced WML compared to alternative presentation methods, instead typically producing equivalent or lower levels of WML across a range of working memory tasks. When higher levels of experienced WML in the HMD-VR condition are found, potential explanations for the finding are offered which are tangential to the use of HMD-VR, and not the direct result of the use of HMD-VR headset itself. It is clear from this review that HMD-VR has application in future working memory research.

Initial indications of what tasks benefit the most from HMD-VR presentation can be gleaned for broader cognitive processes. It is found that papers reporting attention-based tasks, such as continuous performance [169], selective attention tasks [177] and vigilance during a secondary task [173], typically outperform in HMD-VR conditions, presumably where the reduction in external distractions can benefit the most. Similarly, HMD-VR was demonstrated to aid prospective memory, which also requires high levels of attention [174]. BCI tasks did not differ between conditions unless multivariate pattern analysis was used to identify the finer differences between displays [175,176,179]. When isolated to comparisons with SB-VR, HMD-VR did not increase WML when comparing a visuomotor gesture-based task [181], demonstrating how display method and/or subcategory of tasks can affect outcomes.

The papers captured by this review also highlight how the advantages of HMD-VR can be applied in research and application. One such example is utilising virtual elements to compensate for real world limitations. For example, HMD-VR facilitated the easier editing of a mannequin by providing unlimited supplies in Yang et al. [182]. Moreover, HMD-VR was able to present information in a way to enhance working memory processing of low-spatial ability users, providing additional visual information that supported cognitive processing of information [180]. HMD-VR presented VEs also have similar cognitive responses to reality: Peterson et al. [25] found fear responses to walking across beams connecting two high-rise buildings in

the ‘high-VR’ condition, demonstrating how VR can emulate the real world whilst evoking realistic responses to the presented scenario. Providing consideration is given to the differences between real-life and HMD-VR presentations, HMD-VR is still deemed to be suitable in these research contexts.

This review cautions against concluding that, all being balanced, HMD-VR facilitates ‘better’ cognitive processing, nor to discount HMD-VR for a given task, as currently there are too many factors to make such specific judgements. However, considering the range of positive results, explanations for negative findings, and suggestions that HMD-VR does not differ from comparison display methods, this review finds a promising outlook for HMD-VR as a tool in psychological studies.

2.3.3) Combining Head-Mounted Display Virtual Reality and Neurophysiological Recording Methods

The second objective of this review was to examine potential compatibility issues between HMD-VR and neurophysiological methods. Of the 15 gathered papers, 14 papers used EEG in combination with modern HMD-VR, pre-2013 non-modern HMD-VR, and smartphone-HMD-VR. The remaining paper used fNIRS in combination with modern HMD-VR. Each paper demonstrated successful combination of the HMD-VR and neurophysiological recording equipment.

The successful combination of HMD-VR and EEG across headsets is very important for research. Whereas HMD-VR was not anticipated to directly interfere with the BOLD signal captured by fNIRS, different HMD-VR devices have been previously demonstrated to introduce noise to EEG recordings due to the proximity of electrical components to the recording electrodes [56]. Despite this, Harjunen et al. [167] found that HMD-VR conditions had fewer EEG epochs removed due to noise then the SB-VR control. Moreover, Li et al. [177] and Harjunen et al. [167] successfully reported midline electrode results, despite the increased risk of line noise due to the positioning of the cabling of the HMD-VR devices used in their experiments. Papers used lowpass filters of 50Hz and below [16,66,173] or notch filters relative to the

local mains power [169] to successfully remove electrical noise. This builds upon the findings of Cattan et al. [36], who demonstrated that signals emitted from smartphone-HMD-VR did not interfere with EEG signals when combined, and Hertweck et al. [56], who found the noise introduced by certain DB-HMD-VR devices did not interfere with the recorded EEG signal of interest. However, as there are many models of DB-HMD-VR and combinations of smartphones and smartphone-HMD-VR adapters not represented in the present study, care and consideration must still be provided when selecting which equipment to use.

A potential source of noise is the pressure on the scalp when using combined HMD-VR and neurophysiological recording devices, leading to feelings of discomfort in the participant. Perceived discomfort has been linked to experienced cybersickness [168], which in turn can confound results or prematurely end experiments. When coupled with the fact that the use of HMD-VR has previously been found to evoke cybersickness in participants [161,188], additional discomfort from additional head-mounted equipment may worsen the symptoms. Whilst this review does not compare combined HMD-VR and neurophysiological recording methods to HMD-VR alone for cybersickness rates, two papers reported comparative scales of cybersickness between HMD-VR and the alternative presentation method. Juliano et al. [175] reported no significant difference, but Xu et al. [181] found HMD-VR induced higher levels of nausea. Moreover, Harjunen et al. [167] reported that 1 of 12 participants reported feeling nauseous in the HMD-VR condition, though no scale is provided. It is unclear if the low number of studies reporting cybersickness suggests that concerns are minimal, or if it represents an oversight that must be addressed in future studies. However, cybersickness did not prevent results being reported. Thus, whilst this review cannot draw conclusions about HMD-VR and cybersickness, it does not appear combined HMD-VR and neurophysiological methods increase cybersickness to the extent it prohibits combined usage.

It is also found that physical movement, in particular walking, during visuomotor tasks did not prevent successful EEG acquisition during HMD-VR usage [25,178,181]. As HMD-VR is a presentation method that facilitates naturalistic

movement, the ability to replicate real-world actions and navigation is vital for taking full advantage of the display.

This review finds minimal indication that HMD-VR and head-mounted neurophysiological recording methods are not compatible within the configurations examined. There are many successful examples of various categories of HMD-VR being combined with neurophysiological recording methodologies without preventing the successful capture of data. Moreover, HMD-VR induced cybersickness in participants did not prevent the successful completion of the experimental task nor result in the exclusion of data due to noise. The ubiquitous use of head-mounted neurophysiological recording methods is expected, as whilst other neurophysiological recording methods such as fMRI may technically work and have been combined [11,189], they are unexpected to become prevalent as they restrict movement and prevent fully utilising HMD-VR. However, this does not imply there are no compatibility issues between HMD-VR and head-mounted neurophysiological recording devices, as failed combinations would likely not be reported in research articles, and thus would be unlikely to be captured in this review. Therefore, whilst compatibility issues may still arise, this review finds that successful configurations of HMD-VR and neurophysiological recording methodologies which facilitate research are possible, presenting a positive outlook for the future of HMD-VR research.

2.3.4) Shortcomings of the Current Literature, and Recommendations Going Forward

Despite the positive indications of the utility of HMD-VR in research, several shortcomings of the current literature became apparent throughout this review. Based on the captured papers, these problems will be examined, and recommendations given with the aim of improving the standard of future HMD-VR research.

2.3.4.1) Technical Differences Between Head-Mounted Displays and Additional Considerations

This review concludes HMD-VR is an appropriate tool for psychological research, based on findings across a range of modern DB-HMD-VR, non-modern DB-HMD-VR and smartphone-HMD-VR configurations. Some devices have comparable specifications despite being different models, such as the Oculus Rift CV1 and HTC Vive used by Juliano et al. [175] and Li et al. [177] respectively, but offer higher immersion than non-modern HMD-VR (used by Burns & Fairclough [173]) and smartphone-HMD-VR (used by Makransky et al. [16]). It is therefore possible that the differences in immersion contributed to the increased load reported in the 3 papers.

Each paper reporting HMD-VR usage increased experienced WML used either a smartphone-HMD-VR [16] or a 'modern' Oculus DK2 DB-HMD-VR [25,178]. Whilst the low-immersion of smartphone-HMD-VR has been discussed, the DK2 is categorised as a 'modern' DB-HMD-VR device, but is outdated by current standards, possessing inferior resolution, FOV, latency, and overall immersiveness compared to the newer CV1 HMDs. The factors contributing to the relatively low technical specification, specifically FOV and refresh rate, are speculated by Peterson et al. [178] to contribute to the increased WML processing in HMD-VR found in their study. Moreover, the difference in immersion between devices can be seen in the direct comparison conducted by Rupp et al. [6], who found that task performance in CV1-presented VEs outperformed DK2 and smartphone-HMD-VR, attributing the findings to the differences in technical specifications.

Despite the trend of lower-immersion HMD-VR headsets increasing experienced WML, there have been several examples of WML being reduced in conditions using smartphone-HMD-VR [180], non-modern HMD-VR [169], and development-grade DK2 [167,174] configurations. Thus, whilst the HMD used may influence experienced WML, it cannot be the only factor. The most likely explanation is aspects of the HMD-VR configuration which contribute to immersion, but are separate from the specifications of the HMD itself, negatively effecting experienced WML. The effect of these additional aspects may directly influence WML through increasing extraneous

load through task difficulty, or indirectly through cybersickness symptoms or general disorientation. For example, the unintuitive HMD-mounted button used by Makransky et al. [16] does not support sensorimotor perception and introduced additional load by requiring participants to learn to use the input method during the task. Moreover, Peterson et al. [25] tracked participant motion using a Microsoft Kinect motion camera which was used to represent movement in the VE. However, as the Kinect is limited to 30fps, it was speculated that the small delays in movement to representation would be noticeable by participants. Similarly, the visual feed of the webcam mounted to the DK2 used by Peterson et al. [178] to display the real-world was also limited to 30fps, increasing disconnect between the physical action taken and reflection in the VE, thus lowering immersion.

Considering the number of factors which contribute to HMD-VR presented VEs, this review recommends standardising the reporting of details pertaining to the HMD itself and supporting software and hardware used in research. In addition to details about the HMD used including any modifications, information regarding the manufacturer, model, modifications, and supporting hardware such as host computers is important for comparison and replication of research. The reported information must include details on the supporting hardware and VE used, because limitations in these areas can potentially mask shortcomings of experimental designs, or otherwise explain when certain negative effects were found. For example, top-end HMDs limited by sub-par computers could suffer diminished frame rates/latency, reducing immersion and potentially increasing cybersickness [185]. Not only will full reporting of technical specifications lead to a greater understanding of how different HMD-VR devices compare in immersiveness and the impact of working memory processes, it should in turn lead to researchers making informed choices when designing experimental paradigms.

Moreover, there needs to be a greater understanding on the differences between levels of immersion between HMD-VR devices. Whilst this review does not advocate against using smartphone-HMD-VR or low-immersion HMD-VR in research, there is enough evidence to suggest that immersion does play a role in experienced WML.

Whilst using smartphone-HMD-VR instead of DB-HMD-VR is sometimes reported as a potential shortcoming in an experiment [16], papers such as Sun et al. [180] do not discuss how using less immersive smartphone-HMD-VR may influence results relative to more immersive models. This can present a restricted view of the utility of HMD-VR to those unclear on the differences between categories, which can lead to overgeneralisation or misattribution of results to HMD usage generally. Indeed, with consideration between the HMD-VR models available, a second necessary distinction between 'modern' HMD-VR devices becomes apparent. Currently, differences between DB-HMD-VR and HS-HMD-VR have been found within the 'modern' HMD-VR, distinguishing between consumer-grade and high-specification HMD-VR devices. However, there is a clear distinction in the levels of immersion offered between consumer-grade Oculus CV1 and the earlier development-grade Oculus DK2 HMD-VR devices which may lead to differences in cognitive processes when using the devices. Therefore, going forward in this thesis, development-grade DB-HMD-VR devices will be used to refer to early modern models that have comparatively lower specifications than the consumer grade Oculus CV1 or HTC Vive, to highlight the differences between the HMD-VR devices utilised in research. However, it is also anticipated that this expansion will be insufficient going forward, where instead either a generation-based identification (e.g. DK2 is generation 1, CV1 is generation 2, etc.) or a way of categorising immersion may be better suited for distinguishing between future developments.

2.3.4.2) Individual Differences

Another factor which affects the dynamics of HMD-VR and experienced WML is individual differences between participants. Individual differences which have been demonstrated to influence task performance within HMD-VR are often demographic, for example age [190], sex [191], and experience with HMD-adjacent technologies [190]. The clearest effect of individual differences captured in this review is the comparison of low-spatial ability and high-spatial ability groups in Sun et al. [180], who found the former group benefitted from HMD-VR whereas the latter did not. Whilst this example can be explained as the HMD-VR VE provided additional support to the low-spatial ability group, the finding demonstrates how factors that may initially be overlooked can influence experienced load during HMD-VR usage.

This however is a very broad topic with many potential faucets, and what considerations must be given will be largely dependent on the presented task.

The relative novelty of HMD-VR for a participant was identified in several papers as a potential factor that may affect experienced WML, suggesting that new users will experience increased excitement in using the device [180] or towards aspects of a VE [167]. This excitement is speculated to enhance both attention and cognitive processing, thus reducing the experienced WML whilst increasing learning outcomes and other measures of task performance. However, Makransky et al. [16] contests this interpretation by suggesting that excitement can be distracting and will in fact increase load. Regardless, the indications that novelty has an effect means studies which do not consider prior experience, including Makransky et al. [16] and Peterson et al. [25], may unintentionally introduce confounding effects. Similarly, inexperience with input methods, particularly those that are not commonly used by the participant demographics outside of research contexts, could introduce extraneous load whilst participants learn to use the equipment [16]. Therefore, it is recommended that sufficient familiarisation periods with both the HMD-VR-presented VE and input method used, such as those used by Li et al. [177], are provided.

Individual differences affecting experienced WML can also result from the HMD-VR VE, such as experienced presence or embodiment [192]. For example, the effect of experienced presence has been the subject of a review by Coxon et al. [193], who reported that HMD-VR usage increases the feeling of ‘being there’ in susceptible participants. It has been suggested that increased presence results in enhanced focus and reduced distraction [175], therefore participants who are susceptible to the feeling of presence may experience lower levels of WML whilst using HMD-VR devices. Conversely, whilst presence can positively correlate with cognitive load processes (e.g. Li et al. [177]), Makransky et al. [16] reported a negative relationship between presence and results in HMD-VR, finding that whilst presence increased, both the cognitive load experienced and the learning outcomes did not relative to an SB-VR condition.

Individual differences can also influence experienced cybersickness. Levels of reported cybersickness in this review differs between studies, with some finding that symptoms increased in the HMD-VR condition [181], whilst others reporting no difference between presentation methods [175]. Regardless, this difference between experiments, coupled with participants being differently affected by the same stimuli, demonstrates how individual differences influence the experienced symptoms [194,195]. However, it is unclear if these symptoms result from the HMD-VR, the supporting hardware not being sufficient, an aspect of the VE [196], or some combination of these being the cause of the symptoms. It is possible this individual difference can be minimised with research identifying which factors introduce the highest risk when combining HMD-VR and neurophysiological methods, and making appropriate accommodations. For example, if a low frame rate introduces the highest levels of nausea [178], taking steps to increase this through improved specifications or more efficient programming will be important. Moreover, there are common methods employed in VR experiences that are designed to reduce the levels of experienced sickness, such as teleportation-based movement and vignettes during locomotion [197], that can be employed to ease participant where required, dependant on the paradigm used.

Despite the evidence of individual differences interacting with HMD-VR usage, many positive results come from studies that do not consider them, leaving it unclear how important they are. However, the presence of differing findings between groups [180] demonstrates that it is an important factor that should be considered in future research. This review therefore recommends studies should collect and report relevant individual differences in HMD-VR research to provide additional context to results. At minimum, this should include VR experience, experienced presence, and pre- and post-experiment cybersickness scales though this list may expand as more research is conducted.

2.3.4.3) Unbalanced tasks

A recurring trend within the reviewed papers is the separation between tasks in the HMD-VR and comparison conditions. Whilst the same cognitive processes are being

measured between display conditions, there are often differences between how the task is presented or completed that may introduce confounding elements to the experiment or interpretation. Whilst these studies warrant inclusion into this review for demonstrating the research applications of HMD-VR, it is important to understand their limitations and how this may affect the conclusions drawn. For example, the learning materials delivered in Sun et al. [180] were presented in a PowerPoint presentation in real life, opposed to the immersive virtual environment designed specifically to aid understanding the material. One part of the learning materials was examining the Curiosity Mars Rover, which could be presented as a 3D model opposed to a 2D image, allowing more detailed visual inspection in the HMD-VR condition. Such differences could potentially change the underlying cognitive processes, either from processing a more complicated visual representation or potentially having access to more information about the subject, making it unclear to what extent the HMD itself is responsible for differences in experienced load. However, steps can be taken to minimise these differences through careful control of the stimuli, such as ensuring the same information was given in a controlled timeframe identically between conditions, as done by Sun et al. [180].

Yang et al. [182] compared standard pen-and-paper drawing implements in the real-life condition to designing mannequins in 3D space using virtual painting tools in the HMD-VR condition, providing completely different perspectives and way of interacting with the task. Whilst one can conclude that HMD-VR facilitates creativity from the results, the complex HMD-VR experience makes it nearly impossible to determine how much of the HMD-VR device itself was responsible for the reduction of load, and how much resulted from the VE used. Therefore, whilst appropriate for this review as a broad understanding of HMD-VR in cognitive load research, it is important to view these comparisons critically to prevent overgeneralising the findings to HMD-VR usage alone.

Unbalanced conditions were also present in the comparisons with SB-VR, for example Dong et al. [174] used a black-screened slide based prospective memory task in the SB-VR condition where participants must press the appropriate key

depending on the numbers presented. The slide-based task was compared against the HMD-VR condition consisted of a shopping task in a replication of a 3D town, which was navigated using a controller with a rotating chair, and interacted with by reading the numbers aloud and moving their heads to select options on a menu. This comparison between a ‘HMD-VR experience’ to a screen-based task extends far beyond a simple comparison between displays. Whilst Dong et al. [174] does demonstrate an application of HMD-VR, there is an argument that it is comparing two tasks linked only by cognitive process, and thus does not provide much insight into the effect of the headset itself on said processes. This should therefore be portrayed clearly in the publication, to prevent misattributing of the results to being solely resultant from the display method selected opposed to what is presented. Indeed, this criticism extends to all papers that compare against a general ‘VR condition’ without acknowledging the range of differences between the conditions.

Secondary aspects of a paradigm can also unbalance tasks. For example, Makransky et al. [16] faithfully converted a SB-VR environment to HMD-VR but adapted the mouse-based input to the side-mounted button on the HMD itself. As such, the increased experienced WML reported in the HMD-VR condition is suggested to be in part due the novel and unintuitive interaction method. The disparity between input methods highlights two important considerations for researchers: participant familiarity with input methods; and how intuitively the input methods represent an action. Ideally, identical input methods familiar to participants which accurately represent the physical actions taken would be used across conditions. However, balanced motion controls are often not possible as using motion controls on detached 2D screens is fundamentally different from seeing movements accurately represented in fully immersive virtual spaces. HMD-VR can however prevent participants seeing physical keyboards and mouses, which can result in participants misaligning their hands over the keys leading to confusion, increased WML and potentially interrupting the experiment. One potential compromise is using gamepads with similar configurations to VR motion controls, but this would not facilitate many visuomotor tasks nor take full advantage of the HMD-VR’s capabilities. Whilst there is no clear solution present, the input methods selected should be appropriate for the task and remain as consistent as possible

between conditions. The participant should also undergo a familiarisation period prior to the experiment to learn how to use the controls correctly, avoiding artificially increasing WML during the experiment itself [16].

2.3.4.4) Unexplored Comparisons

Whilst this review succeeds in broadly comparing HMD-VR to other display methods, no captured neurophysiological paper compares HMD-VR to CAVE or augmented reality (AR). The lack of CAVE and AR representation is an important gap in the literature, as both offer an alternative “middle ground” between SB-VR and real-life currently only explored using behavioural measures (e.g. Demitriadou et al. [198] and Halabi et al. [199]). The lack of comparisons with CAVE is not surprising considering the practical limitations [27], however it would allow VEs to be explored whilst maintaining one’s physical body as opposed to virtual avatars. Conversely, AR overlays virtual elements onto the real world, either directly using ’see-through’ HMDs (used by Funk et al. [200] and Werrlich et al. [201]), or onscreen using mobile devices with cameras (used by Huang et al. [202]). There is an argument that could be made that Peterson et al. [178] utilised a form of AR, as the real world was presented virtually, however this does not necessarily fit the earlier description of overlaying virtual elements of the real world. For the purpose of this review, Peterson et al. [178] is considered VR the same way a recording of a roller-coaster presented within HMD-VR is.

No comparisons with HS-HMD-VR and an alternative display were found. The findings suggest that level of immersion has some effect on experienced WML, however devices that offer the highest level of immersion are not represented in the current literature. This is an important oversight for two reasons, firstly because it is currently unknown how the highest level of immersion influences experienced working memory processes compared to alternative displays. Secondly, it could potentially indicate that HS-HMD-VR is not suitable for EEG research. EEG has been combined with HS-HMD-VR, and found to have limited impact on recorded frequency bands [56], however there may be practical limitations preventing HS-HMD-VR being deployed in working memory paradigms. Additional research

focusing on HS-HMD-VR in working memory paradigms is therefore required, especially considering the rate at which HMD-VR technology is advancing.

All but one experiment used EEG, and whilst HMD-VR has been successfully combined with methods such as fNIRS [203], little can be drawn about BOLD responses changes in the brain during HMD-VR usage. It is therefore possible this review did not capture examples of HMD-VR interfering with non-electrophysical neurophysiological recordings, or potentially missed aspects of the neurological response not detectable with EEG. Thus, there may be important considerations regarding combined methodologies that will be uncovered by future research.

2.4) Conclusions

This review systematically examined combined use of HMD-VR and neurophysiological recording methodology in the study of working memory processes. A range of working memory tasks including attention, learning, and BCI, are represented in the current combined HMD-VR and EEG literature. The results reviewed indicate that the use of HMD-VR does not inherently negatively influence experienced WML across a range of cognitive tasks, instead finding no difference between displays or lowering the measured WML performance. Instances where increased WML was reported in the HMD-VR condition could be explained by secondary factors of the VR configurations used. The findings suggests that HMD-VR technology is suitable for wider use in psychological and neurological research, and potential real-world application. Moreover, it was found that smartphone-HMD-VR, non-modern DB-HMD-VR, development-grade modern DB-HMD-VR, and consumer-grade modern HMD-VR have been successfully combined with EEG methodologies to acquire neurophysiological measures of working memory. An example of combined modern development-grade HMD-VR and fNIRS was also found, demonstrating how HMD-VR is being examined using other neurophysiological recording methodologies. However, current research using modern HMD-VR is a relatively novel field, and research comparing HMD-VR to alternative displays across a wider range of working memory processes, particularly

using HS-HMD-VR devices, is required to understand how the technology can be best applied in research and application.

Chapter 3) Event-Related Potentials in Response to an Arithmetic Working Memory Task Presented in a High-Specification Head-mounted Display Virtual Reality Virtual Environment.

3.1) Introduction

3.1.1) High-Specification Head-Mounted Display Virtual Reality in Neuroscience Research

The systematic review conducted in Chapter 2 found that EEG has been successfully used to record neurophysiological measures of experienced WML in combination with a range of HMD-VR devices [16,167]. WML-related ERPs have been acquired using Google Cardboard smartphone-HMD-VR [204], development-grade Oculus DK1 and DK2 DB-HMD-VR systems [167,176,178], and the consumer-grade HTC Vive DB-HMD-VR [66]. Combined HMD-VR and EEG has further been reported in published research outside of the systematic review, including Cattan et al. [64] using smartphone-HMD-VR in a P300-BCI task and Aksoy et al. [205] using a consumer-grade HTC Vive in an N-back task. The use of the combined EEG and HMD-VR extends past WML research, with Stolz et al. [206] reporting successful ERP acquisition in an HMD-VR experiment studying responses to face presentation using the consumer-grade Oculus CV1.

Although the successful combinations of HMD-VR and EEG demonstrate the suitability of using both together in neuroscience research, lower-immersion HMDs may contribute to unwanted extraneous load in WML tasks and other confounding factors in research. For example, it has been demonstrated that the lower resolution of low-immersion HMDs makes visually navigating a VE more difficult, resulting in lower learning outcomes and engagement [6] or leading to increased cybersickness symptoms relative to higher-immersion devices [207]. Secondary aspects of the low-immersion HMD-VR configuration can also inhibit working memory performance, for example if the HMD-VR utilises unintuitive or confusing input methods [16]. However,

no paper captured by the systematic review utilises HS-HMD-VR devices such as the HTC Vive Pro despite offering the highest levels of immersion, highlighting a major gap in the current literature.

Many low-immersion development HMDs have been discontinued in favour of higher-specification models, such as the consumer-grade HTC Vive [66,182] or Oculus CV1 [179,181]. Relative to the development-grade HMDs, the consumer-grade devices offer larger resolution displays, enhanced internal components, and other improvements making them an attractive option for research applications. Moreover, these consumer-grade HMDs have been surpassed by HS-HMD-VR devices which offer the highest levels of immersion, and as such are gaining interest for use in neuroscience research [208]. However, the wide range of available HMD-VR devices introduces a large disparity in what 'HMD-VR' can be in research. For example, the development-grade Oculus DK2 is a very different headset compared to high-specification Vive Pro (Figure 3.1). The DK2 contains a singular 5.7" OLED screen with a combined 1920x1080 pixel resolution,100-degree field of view, no integrated audio, and weighing 440g. The straps are elasticated fabric that are tightened around the head, which guide the data and power cables over the midline of the scalp. In comparison, the Vive Pro has two 90Hz AMOLED 3.5" screens with 1440x1600 pixels per eye screen with a combined 110-degree field of view, providing a clearer, higher resolution presentation of the VE and more of the periphery compared to the DK2. The Vive Pro uses a ratchet-style tightening method of affixing the HMD to the head, being held in place from the cushioned face pad and a cushioned pad at the back of the head, in combination with an overhead Velcro midline strap. Unlike the DK2, the side straps do not make contact with the head. The Vive Pro trails the power and data cables down the left side of the head. However, the improvements to the Vive Pro also increased the weight to 555g, ~25% over the DK2.

(a) Front View

(b) Side View

(c) Top View

Figure 3.1: *Annotated side-by-side images of an Oculus DK2 (left) and HTC Vive Pro (right) from the front (a), side (b) and top (c). The similar sizes of the main units are seen in all 3 images. The interpupillary distance and lens distance calibrations for the Vive Pro are highlighted in images (a) and (b). The integrated hard plastic straps and cushions on the Vive Pro and elastic straps and face cushion of the DK2 are highlighted in images (b) and (c). The data trailing wires in the centre strap of the Oculus DK2 and on the left side (when worn) are highlighted in orange image (c).*

Whilst most modern HMD-VR devices facilitate some form of audio delivery, HS-HMD-VR devices are capable of efficiently combining the higher-fidelity visual presentation with higher-grade auditory hardware to facilitate immersive VEs. The Vive Pro has cushioned on-ear headphones which are integrated directly into the device, allowing directional sound to be delivered whilst blocking out a degree of external noise. In comparison, the Oculus Rift CV1 has integrated off-ear headphones which do not block external noise; the HTC Vive has a headphone jack to use external headphones, but therefore requires an additional wire to be used which can potentially obstruct a participant's movement; and the Oculus DK2 has no in-built audio integration and relies on computer speakers or external headphones. The increased immersion from the combined visual and auditory elements offered by HS-HMD-VR has implications for experimental paradigms which utilise speech or otherwise emulate real world situations by increasing presence [209].

The technical advancements of the Vive Pro over other headsets could be expected to promote WML performance [6], however there are several aspects of the HMD's design that could potentially inhibit EEG data acquisition. Firstly, the enhanced screen and components could introduce increased electrical noise to the EEG recording due to close proximity to the recording electrodes. Whilst electrical noise originating from the power cable (50/60Hz) or the screen (90Hz) should be identifiable using frequency extraction and eliminated by standard EEG filtering techniques, lower frequency noise within the 0-40Hz range may prevent accurate EEG recordings of frequency bands commonly used within neuroscience research [210,211]. Secondly, the left-mounted power cable may introduce line noise with a left hemisphere bias. Thirdly, the design of the head straps, particularly on the back of the head, may block or otherwise move the electrodes on the scalp during the tightening-up process. This may also prove uncomfortable for participants using EEG headcaps with protruding electrode holders, such as the SPESMedica Softcap.

There have been examples of combined HS-HMD-VR, specifically the Vive Pro, and EEG. For example, Hertweck et al. [56] examined the noise produced by the Vive Pro using time-frequency EEG analysis, and reported that the HMD only introduces

the easily filterable 50Hz line noise to the recording or a 100Hz frequency, above what is typically analysed in EEG experiments. However, some experiments using combined Vive Pro and EEG do not employ a standard EEG array integrated into a cap, instead using a modified HMD-VR faceplate (Figure 3.2) to apply electrodes to the prefrontal and frontal areas only [73,212,213]. Furthermore, WML studies using the Vive Pro with whole-head arrays do not include event-related potential (ERP) in the analysis, with successful combinations only utilising time-frequency analyses [56,214]. As demonstrated by the systematic review, many studies utilise ERP measures of WML, thus it is important that HS-HMD-VR is compatible with the analysis method for the method's application in research. Moreover, the effect of different levels of induced load within HS-HMD-VR is underexplored in the current literature, rendering it relatively unknown if the use of such devices interacts with different levels of experienced load. Therefore, before it can be understood how to best take advantage of HMD-VR in research, it must be confirmed that higher specification HMD devices can be used in combination with whole-cap EEG recording equipment during an ERP recording.

Figure 3.2: *Image of a modified Vive Pro with the Looxid Link Mask and EEG recording system, as used by Costa et al. [212]. This configuration uses dry AF3, AF4, AF7, AF8, Fp1, Fp2 electrodes, and uses FPz as the reference.*

Moreover, auditory and visual stimuli are typically combined into a singular experience during HMD-VR research, presenting a unified VE to the participants. However, there is little if any research investigating the presentation modalities separately in the context of HMD-VR induced WML. The visual and auditory presentations of the same information will separately evoke the visuospatial sketchpad and phonological loop components of the multicomponent working memory system respectively [215], which in turn could result in different levels of experienced WML despite the same information being processed [216,217]. Indeed, studies comparing the presentation of visual and auditory stimuli using HS-HMD-VR compare visual-only presentations to combined audio-visual, instead of auditory-only presentation [218]. It is therefore difficult to isolate the effect that the HMD-VR audio hardware has on the EEG recording.

A comparison of ERP EEG recordings taken during a working memory task presented using HS-HMD-VR would build upon the outcomes of the systematic review. Firstly, recording ERP EEG signals during a task presented using HS-HMD-VR would examine if the techniques can be combined in an ERP experiment, or if the combined use impacts the electrophysiological recording and prevents data acquisition. Secondly, it would further explore the dynamics between HMD-VR usage and levels of experienced load by utilising different levels of load within HMD-VR presentation, opposed to the singular levels of load commonly employed in the studies identified by the systematic review.

3.1.2) Mental Arithmetic

There are many WML paradigms that could take advantage of the intricate perceptual experience HMD-VR facilitates, such as learning outcome comparisons in virtual laboratories [16] or performance in realistic vehicle simulations [32]. However, mental arithmetic tasks provide a versatile method of inducing different levels of WML currently unutilised in HMD-VR research. Mental arithmetic calculations are complex cognitive processes linked to the central executive, visuospatial sketchpad and phonological loop working memory components [219,220]. Each equation is solved within working memory over several stages: the equation is encoded to WM;

the appropriate operation (addition, multiplication, etc.) is retrieved from long term memory; each intermediate step is calculated and maintained; and a response is provided [219]. Arithmetic paradigms in research are divided between ‘production’ tasks, where the answer must be calculated and reported; and ‘verification’ tasks, where potential answers are presented, and the participant must determine which if any are correct. Verification tasks may present a single potential solution which is identified as correct or incorrect, or multiple which must be chosen between. Arithmetic paradigms can also differ on how the equation is presented, for example displaying operands (the numbers) and operations together or in sequence [204,221,222].

Mental arithmetic tasks offer several major benefits for HMD-VR research using WML paradigms. Primarily, the same or equivalent arithmetic stimuli can be presented visually or auditorily [223], allowing for different aspects of the HMD-VR experience and the respective working memory component to be examined. Both visual and auditory presentation of working memory tasks have been used in working memory paradigms [224,225] to probe the visuospatial sketchpad and phonological loop of the multicomponent working memory model respectively, however the control over task difficulty offered by arithmetic stimuli allows for levels of load to be manipulated across modalities within the same experiment. Therefore, even if visual and auditory arithmetic questions do not evoke equivalent loads [217], it will allow exploration of how the visual and auditory aspects of the HMD-VR experience are processed differently within the brain, and by extension how said processes are effected by different levels of WML. Moreover, as only the arithmetic question needs to be presented, simple VEs can be utilised which avoid potentially distracting or confusing elements used in more complex learning tasks.

A mental arithmetic paradigm is ideal for the current WML-based exploration of HMD-VR for the level of control it offers over the level of induced load. For example, it has been long understood that increasing the problem size (the intrinsic difficulty of an arithmetic task) of an equation through increasing the number of digits and carry overs positively correlates with errors made [226], indicating higher levels of WML.

Moreover, the problem size of addition questions can be empirically controlled and categorised using the ‘Q-value’ calculation [227] (Figure 3.3). This calculation accounts for the steps taken to correctly solve an equation, including carry-overs, by summing the logarithms of each pair of digits at each space of the equation (units with units, tens with tens, etc.) with the total of the digits in a given position within the equation. Each step of solving arithmetic questions has linked to WML [219], and it has been found that increasing the number of steps, for example the number of carry overs [228], increases experienced load. Arithmetic problems can then be separated into distinct brackets of Q-values, allowing questions to be grouped into categories ranging between ‘easy’ to ‘too difficult’ [229].

a: $Q(x_1 + y_1) = \log[x_1 + y_1 + (x_1 + y_1)]$

b: $Q(x_1 + y_1) = \log[x_1 + y_1 + (x_1 + y_1) + 10 + (x_1 + y_1 - 10)]$

c: $Q(x_1x_2 + y_1y_2) = Q(x_1 + y_1) + Q(x_2 + y_2 + c)$
$=\log[x_1 + y_1 + (x_1 + y_1) + 10 + (x_1 + y_1 - 10)] + \log[x_2 + y_2 + (x_2 + y_2) + c_2 + (x_2 + y_2 + c_2)]$

Figure 3.3: *The Q-Value Equations. These equations reproduced from Spüler et al. [227] are the calculations used to determine the difficulty of various addition equations. Equation (a) is used for single-digit plus single-digit equations with no carry overs, (b) is used for single-digit plus single-digit equations including carry overs, and (c) for multi-digit plus multi-digit equations including a single carry over. These equations can be expanded as required to accommodate sums with multiple digits.*

Moreover, current research illustrates how different aspects of mental arithmetic paradigms or problems can influence experienced WML [219,230,231]. For example, it has been found that limiting presentation times of the arithmetic question during calculation increases experienced WML [228]. Moreover, how the question is presented affects the experienced load, for example auditorily presented equations are found to induce higher WML than visually presented equivalents [217]. Within visual presentation, Blankenberger [232] found presenting the information using a numerical format (e.g. ‘2+4’) induces a lower load than word-based presentation (‘two plus four’), which can be attributed to the additional step of converting the

words to numbers. Conversely, increased WML from factors outside of the arithmetic question itself results in decreased arithmetic performance [233]. For example, Imbo et al. [234] found that participants utilised inferior strategies for solving arithmetic problems when under increased WML. Based on these findings, it is possible to design a paradigm that avoids introducing extraneous load from the question presentation within the HMD-VR VE.

3.1.3) Mental Arithmetic in Event Related Potential Research

Mental arithmetic tasks are commonly employed in EEG ERP research, which typically examine ERP responses to equation presentation in production tasks and solution presentation in verification tasks. A common ERP response in the arithmetic literature is the P300 response [235–237], and can be detected across the brain in response to arithmetic stimuli [236]. The P300 has been linked to several cognitive processes associated with different aspects of mental arithmetic calculations, including the processing of a presented arithmetic question [204,222,238], evaluation of task difficulty during question presentation [235], and target detection of presented solutions in verification tasks [237,239]. Typically, a larger amplitude P300 is associated with a lower WML in arithmetic contexts for both question and solution presentations [204,223,238,239]. The location of the largest P300 response is also modulated by the difficulty of the arithmetic task and the modality it is delivered in [131], but are typically reported in the parietal regions [204,236,240] and central electrodes [204,223]. The hemisphere the largest P300 response is found is inconstant, with studies reporting the largest responses originate in the left hemisphere [237], right hemisphere [241] and the midline electrodes [223,237]. This results from the specific arithmetic operation and modality utilised, as Dickson & Federmeier [223] demonstrated responses were larger in the left parietal lobe for visually presented and easy questions, and larger in the right parietal lobe for auditory and hard questions.

In addition to the P300, earlier N170 components have been reported in response to arithmetic stimuli [241]. N170s are more commonly associated with recognition of faces [242], and have been demonstrated to be sensitive to different levels of WML

within face-perception tasks [135]. However, N170 components have been shown to be sensitive to the problem size of arithmetic questions. For example, Moore et al. [241] and He et al. [243] found that more difficult arithmetic tasks increased the N170 response in the left hemisphere over the right and medial regions, reported to originate from the parietal lobe and fusiform gyrus respectively. These early N170s are speculated to be associated with the encoding of the arithmetic stimuli [241,243], however this has currently only been reported in verification tasks, therefore their presence in production tasks is unknown.

Post-P300 slow-wave components (SWCs) have also previously been reported in arithmetic paradigms, typically manifesting as either a frontal negative response, parietal positive response, or combination of these within the 450-750ms range [131,244]. Arithmetic SWC are linked to several cognitive processes depending on the paradigm employed. For example, arithmetic SWCs represent the identification of ‘rule violations’ in arithmetic verification tasks, with presented correct responses (lower-WML) having smaller responses than presented incorrect responses (higher-WML) [223,240,245–247]. In production tasks, the arithmetic SWC is associated with the mental calculation process [131,248] required to arrive at the answer. Most commonly, SWCs are characterised by the positive inflections in the parietal [240,244] and parietal-central regions [221], and is sometimes referred to as a positive SWC or the late positive component (LPC). Similar to the P300, positive SWCs are reported across hemispheres, being found to be of the largest amplitude in the right-hemisphere [223,245], left-parietal region [236], within the midline electrodes [237], or independent of any laterality [240]. The amplitudes of SWCs have been found to increase with WML within arithmetic verification tasks [223,240,246] but have also been reported in response to the presentation of increasingly difficult arithmetic questions in response to both visual and auditory questions [248].

3.1.4) Aims of the Present Study

The technical specifications of HMD-VR devices available in neuroscience and WML research have increased drastically since modern HMD-VR devices became more

widely accessible, with high-specification HMD-VR providing the greatest levels of immersion through the large resolution screens and audio delivery methods. However, high-specification devices such as the HTC Vive Pro have not been utilised in combination with EEG techniques in WML-based neuroscience studies, and it is currently uncertain if HS-HMD-VR devices can be successfully used when acquiring ERP responses. This study aims to acquire N170, P300 and SWC ERP responses from a working memory mental arithmetic task presented visually and auditorily using HS-HMD-VR combined with EEG. The arithmetic task will present a series of addition questions comprising of single-digit or double-digit numbers to evoke distinct levels of WML within participants. Questions will be presented either visually or auditorily within the HMD-VR VE, and participants will be presented with each variation of presentation modality and question difficulty. This aim will be achieved through identifying ERP responses to visual and auditory mental arithmetic stimuli presented using an unmodified HS-HMD-VR device, and recorded using a standard EEG headcap. The ERP responses within selected component time windows will then be compared based on difficulty/load, electrode location and hemispheric location using a repeated measures design.

3.2) Methods

3.2.1) Participants

This study recruited 22 university students (Mean Age ± S.E.M = 21.36 ± 0.6 years old, range: 18-27 years old), 12 males (M= 22.4 ± 0.9 years old, range: 18-27 years old) and 10 females (M= 20.1 ± 0.7 years old, range: 18-24 years old), with normal-to-corrected vision through opportunity sampling from the Hull University and University of York campuses. The number of participants recruited within EEG studies typically ranges between 10-30 and averages around 20 (see Larson & Carbine, [183] and Clayson et al. [249]), thus the number of participants is standard within EEG/ERP research. Two participants were left-handed (1 male, 1 female) based on responses to a modified version of the Edinburgh Handedness Inventory [250]. Every participant completed a paper-based health questionnaire to exclude any with pre-existing conditions or who are taking medication which could affect cognitive function. The exclusion criteria were predefined by the researcher, and

included neurological conditions such as epilepsy or conditions requiring an electrical stimulator (such as a pacemaker) which may introduce noise to the EEG recording. Fourteen participants had no prior experience with HMD-VR. Of the 8 who had experience, 5 had less than 1 hour of exposure, and 3 had 1-10 hours of use. One participant was excluded from data analysis for being unable to finish the experiment due to cybersickness. The study was conducted in conformity with the Declaration of Helsinki [251], received local ethical approval from the Hull York Medical School Ethics Committee (Reference 1303), and all participants provided fully informed consent before beginning the experiment.

3.2.2) Materials and Apparatus

3.2.2.1) Hardware and Software

The experimental stimuli were constructed and presented using PsychoPy [55], which also captured the behavioural data. The audio questions were generated using Python [252] and the Google Text-To-Speech [253] plugin, which were then converted from .MP3 to .ogg file formats using the MediaHuman Audio Converter [254] to be compatible with PsychoPy. The post-test SSQ [160] was also constructed in PsychoPy and presented within the VE for participants who opted to complete it before taking off the HMD.

The stimuli presentation, virtual environment and behavioural data recording was managed by a desktop computer using an i5-2400 processor and Nvidia GTX 980 GPU, and presented using an HTC Vive Pro headset. A wire connecting the headset to the computer was trailed behind the participant to prevent obstructing answer input. The Vive Pro has integrated headphones that were placed over the ears, and adjustable straps for comfort. The HMD-VR hardware was managed by SteamVR software. The HMD was calibrated for each participant by adjusting the integrated straps, the FOV adjuster and IPD knob to ensure the HMD was properly configured for each participant.

The Vive Pro presented the VE in a 1440 x 1600-pixel resolution, and the stimuli was rendered on a cinema-style virtual screen within the VE at 1920x1080 pixels using the BigScreen Beta application. The distance from the virtual screen to where the participant were located within the VE was kept consistent by recentring position of the participants before the onset of the experiment, and the entire screen was within the participants view when facing forwards. Participants interacted with the VE through the Vive Wand controllers, which were represented as cartoon-style hands within the VE. These wands emulated a mouse when pointing at the virtual screen, and clicks were emulated through pulling the trigger. Within the VE, the 'hands' emitted a laser-pointer based on the location and orientation of the wand, allowing participants to accurately input responses. The headset and input devices were tracked by two Vive 'Lighthouse' sensors placed on tripod stands positioned to provide complete coverage to the experimental area. Ceiling-affixed retractable wire guides were employed to relieve downwards pressure on the back of the head from the HMD-VR data cable connecting to the control unit (Figure 3.4).

(a) (b)

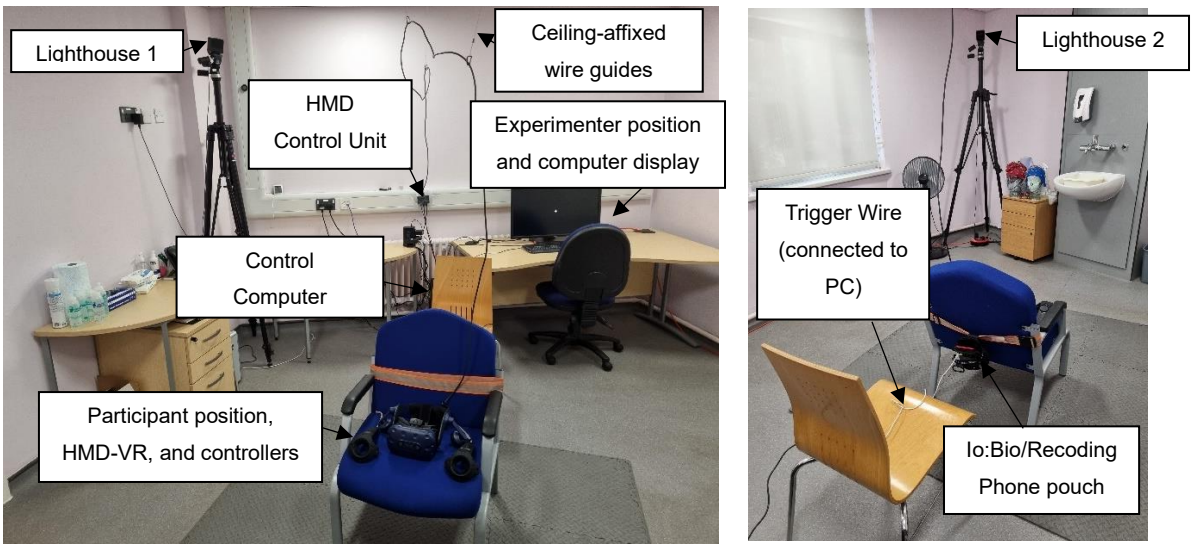


Figure 3.4: *Annotated images showing the configuration of the laboratory during the experimental procedure facing the participant (a), and from behind the participant (b). The major components of the VR setup are shown in image a, whilst the EEG recording device is shown in image b. The virtual screen was configured to appear in front of the participant based on the position of the chair.*

3.2.2.2) Electroencephalography Recording

Electroencephalography signals were recorded using 19 channel (FP1, FP2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2) tin-electrode Spes Medica Sleepcaps, using the 10-20 configuration and a physically linked-ear reference. When required, an additional headcap net was placed over the EEG cap to improve the connection between the electrodes and the scalp. The headcap was connected to an Io:Bio EEG device [255], recording at a 250Hz sampling rate which was then transmitted via Wi-Fi to an Asus Zenphone 6 mobile phone recording device. The Io:Bio device was also connected to the computer via a proprietary cable to receive EEG event triggers.

3.2.2.3) Mental Arithmetic Task

The mental arithmetic paradigm used in this current experiment uses a production task with the simultaneous presentation of the entire arithmetic equation (except the solution), based on the studies conducted by Moore et al [241], Jasinski & Coch [237], and Dickson and Federmeier [223] and adapted for HMD-VR presentation. Simultaneous presentation of the entire equation was selected as it provides a naturalistic way that arithmetic questions are presented in the learning environments, and to prevent flashing stimuli that may introduce visual discomfort. A production task was selected as it allowed for the motion-based controllers of the HMD-VR configuration to be utilised in a naturalistic way for inputting information into a virtual number pad (the sequential pressing of multiple buttons), as opposed to selecting singular options or responding verbally when verifying a potential answer.

Participants were presented arithmetic questions either visually or auditorily within an HMD-VR environment. Two distinct difficulties of arithmetic questions were created based on the Q-values [227] and categories used by Chin et al. [229], comprising of ‘easy’ single-digit plus single-digit and ‘hard’ two-digit plus two-digit questions. The ‘easy’ questions were selected by generating every combination of single-digit questions, then selecting 60 of the 81 at random using Microsoft Excel. This avoided a middle-slice bias of Q-values which were centred around the digit 5. Sixty questions were selected to remain consistent with previously published literature

[256]. No Q-value in this group exceeded 2 (min=0.602, max=1.708, mean=1.365, SD=0.269), consistent with the ‘easy’ difficulty category used by Spüler et al. [227] and Chin et al. [229]. For the ‘hard’ questions, 81 two-digit questions were generated using the =RANDBETWEEN(10, 99) function in Microsoft Excel to generate 2 two-digit numbers for each potential question. The equations were then ordered on their resulting Q-Values from lowest to highest, and the first 60 with a Q value greater than 2.3 were extracted, ensuring there was a gap in difficulty between the highest-Q ‘easy’ question and the lowest-Q ‘hard’ question. No Q-value exceeded 4 (min=2.33, max=3.352, mean=2.914, SD=0.317), which is defined as ‘too difficult’ in the categories used by Spüler et al. [227] and Chin et al. [229].

These questions were randomly assigned to 12 blocks of 10 based on difficulty to prevent clustering of close Q-value questions. Due to the limited number of possible combinations of single-digit equations and to ensure conditions were balanced, blocks were repeated between visual and auditory presentations, totalling 24 blocks across 4 conditions: ‘Easy Visual’; ‘Hard Visual’; ‘Easy Auditory’; and ‘Hard Auditory’. The order of these blocks was pseudo-random, with the order of the blocks randomised then modified to prevent two blocks of the same condition being presented successively. The order of the blocks was consistent between participants, though the exact order of the questions within the blocks were randomised per participant.

The edges of the virtual screen within the VE were rendered invisible by filling the background of the paradigm as solid black and presenting it in the ‘void’ environment in BigScreen (Figure 3.5), which consists of a full black environment with no other visual or auditory elements. This gives the illusion at each stage of the experiment that the stimuli or response inputs appear as if they are floating in space before the participant.

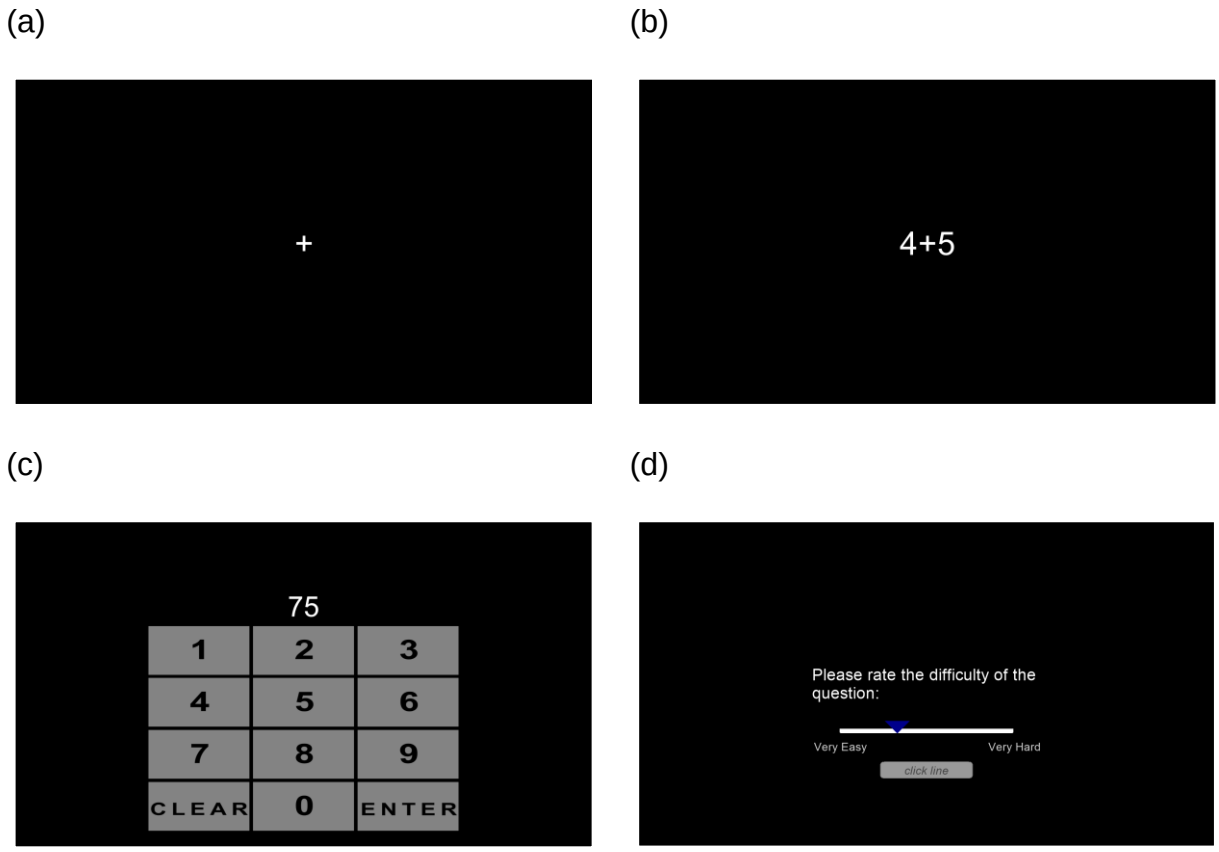


Figure 3.5: *The 4 major screens of the experiment presented using the HMD-VR BigScreen program: The fixation cross (a), an example of a visual question (b), the response input screen (c), and the subjective difficulty rating (d). The fixation cross is also used when auditory questions are delivered to the participant.*

3.2.2.4) Behavioural Measures and Participant Demographic Information

Prior to the onset of the arithmetic task, two factors relevant to prior HMD-VR experience were collected in the preliminary questionnaire (Appendix 1). Each participant reported prior experience using HMD-VR on one of five time durations: “Never”; “< 1 Hour”; “1 Hour to 10 Hours”; “10 to 24 Hours”; and “> 24 Hours”. Handedness was captured via a modified Edinburgh Handedness Inventory [250]. Participant response to the Edinburgh Handedness Inventory did not dictate which hand they used to input answers, as controllers for both hands were provided to prevent technical issues.

Three behavioural measures were recorded for each maths question: response time; response accuracy; and subjective ratings of question difficulty. Response time is automatically captured from the onset of the response input until the ‘enter’ button is

pressed. Response accuracy and distance from the correct answer were also automatically recorded when the “enter” button is pressed depending on the correctness of the participant’s response. After each response, participants were asked to ‘Please rate the difficulty of the question:’ using the subjective difficulty rating scale. The subjective difficulty scales are based on those used by Paas and colleagues [257–259], specifically, modifying the 7-point “extremely easy” to “extremely difficult” scale used in Ayres’ [260] arithmetic task for use in HMD-VR. A continuous scale is simulated within the HMD-VR VE by using a 500-point scale, which facilitates response input when using HMD-VR motion controllers by allowing any point in the scale to be selected. The minimum and maximum ends of the used scale is marked as “Very Easy” and “Very Hard” respectively, and no numbers or sub-categories were presented to prevent visual overload or confusion on the limited screen space.

The 16-item SSQ (Appendix 2) was presented at the end of the arithmetic task to measure symptoms associated with cybersickness. Participants were given the choice to complete this within the HMD-VR VE or using a paper version. Participants who ceased the experiment early were only given the paper version option and asked to remember how they felt prior to removing the HMD.

3.2.3) Experimental Design

The experiment design used in the current study comprises of two separate within-subject 2-way comparisons between arithmetic question difficulty within presentation modalities. Equation difficulty is manipulated by using single-digit (‘easy’) and double-digit (‘hard’) addition questions which fall within defined q-value ranges (defined in section 3.2.2.3). Comparisons are conducted within the visual and auditory presentation modalities. One presentation modality is visual, and the other presentation modality is auditory, but no comparisons between presentation modalities are conducted.

3.2.4) Experimental Procedures

Upon arriving in the EEG laboratory, participants were provided with an information sheet (Appendix 3) and an informed consent form (Appendix 4). Upon reading and signing the consent form, participants were given the health questionnaire (Appendix 5) and the preliminary questionnaire to complete. An overview of the task was explained to the participant, followed by a demonstration of how to operate the Vive wands.

When ready, the researchers put the EEG headcap onto the participant, and the Vive Pro was placed over the EEG cap (Figure 3.6). The participant was able to alter the interpupillary distance and focus within the Vive Pro so that the VE was clear. This was confirmed by having the participant read aloud presented text introducing and explaining the experiment. The participant was provided with a Vive Wand for each hand, and instructed to use whichever they felt was comfortable for them. This also ensured that both hands were represented in the virtual space to reduce visual-proprioception disconnect. As only one controller could be used to interact with the virtual screen at a time, participants were instructed to ‘activate’ the controller in their preferred hand by pulling the trigger, which also functioned as a ‘mouse click’ on the virtual screen. Controller activation was confirmed by the appearance of a laser-pointer style beam originating from the controller when pointing at the virtual screen within the VE. Whilst it is possible to ‘activate’ the unused controller through an additional trigger pull, doing so ‘deactivates’ the other controller, thus participants were instructed to avoid doing so during the experiment as to prevent extra button presses, participant confusion, and potential EEG noise.

(a) (b)

Figure 3.6: *A view of the combined HMD-VR (Vive Pro) and EEG headcap on a participant from the front (a) and side (b) views. The chinstrap kept the EEG cap in place. The majority of the EEG electrodes do not make contact with the HMD-VR device. The FP1, FP2, F7, and F8 electrodes (Orange) make contact with the frontal face cushion. The Fz, Cz, Pz, and ground electrodes (Green) make contact with the central Velcro head strap. The O1 and O2 electrodes (blue) make contact with the head cushion at the back of the HMD.*

Next, the participant underwent a 10-question training session consisting of 5 visual and 5 auditory arithmetic questions to familiarise themselves with the procedure, controls, and ensure the volume was at a comfortable level. These questions were selected from the pool of unused questions to prevent repetition. Upon completion of the training phase, the participant was given the opportunity to ask any questions and reminded to answer each arithmetic question as quickly and accurately as possible. The lights were dimmed in the lab to minimise external distractions, and the EEG recording commenced.

Each trial began with a 1000ms fixation cross (Figure 3.7a) followed by a 5000ms stimulus presentation (Figure 3.7b). In the visual condition, the question remained on screen for the 5000ms period, with the fixation cross acting as the plus sign. In the auditory condition, the fixation cross remained for the 5000ms period whilst the audio file played. Participants were instructed to rest their forearms on the chair's armrest during question presentation to avoid movement artifacts. The participant was then presented with a number-pad (Figure 3.7c) to which they physically pointed the wand controller at to input their responses. Each number was selected individually, with the

'enter' button submitting the response, and the 'clear' button resetting the number in the event a mistake was made. Upon submitting their response, the subjective difficulty rating for the question was presented (Figure 3.7d). The participants were provided with unlimited time for these inputs. In total 240 arithmetic questions were presented to the participant. Event labels for the EEG triggers were generated based on condition and participant accuracy.

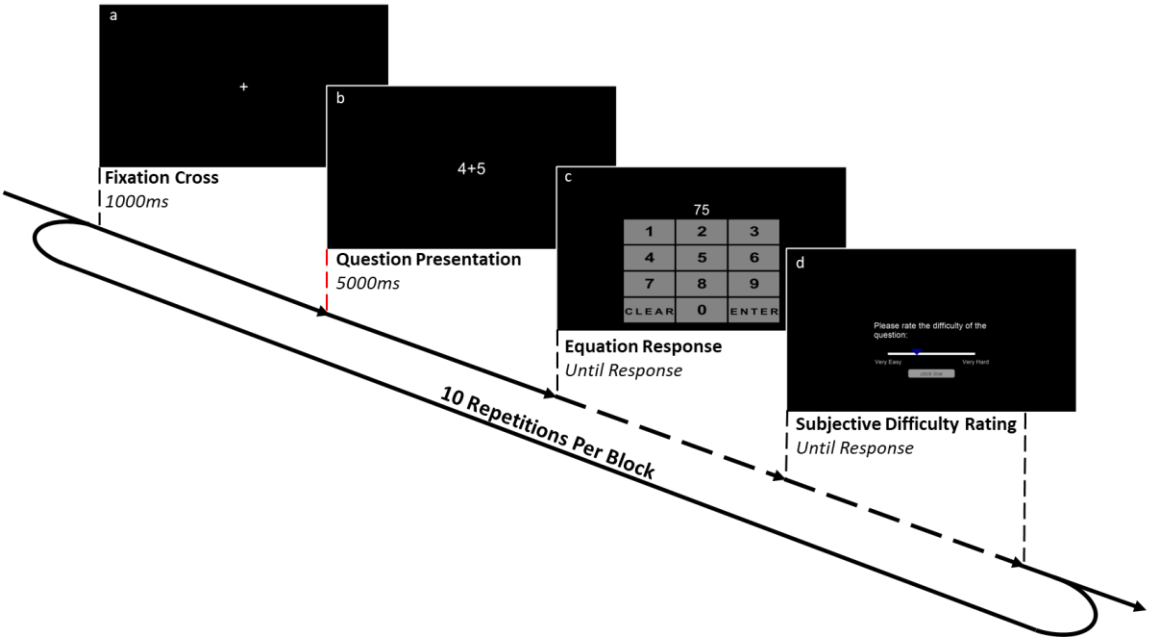


Figure 3.7: *Example of a visually presented addition question trial and block. This diagram shows the process of each trial presenting an arithmetic question. a: Fixation Cross: A fixation cross is presented for 1000ms in the centre of the screen to direct participant's attention. b: Question Presentation: An example of an easy-visual condition. In the auditory condition, the fixation cross remains as with Figure 3.7a, and the equation is read aloud. This stage lasts 5000ms and includes an ERP trigger at the start of the stimuli presentation (represented by the dashed vertical red line on the diagram). c: Response Input/Equation Response: A number pad is presented which is interacted with by pointing the controller at the number to select the response, with no time limit and the ability to delete mistaken inputs. d: Subjective Difficulty Rating: The 500-point linear scale from 'Very Easy' to 'Very Hard'. The blue arrow represents the currently selected point.*

Upon completion of the EEG/arithmetic study, the participant was presented with the SSQ. When complete, each participant was provided a debriefing form (Appendix 6) and given the opportunity to discuss the experiment before leaving the laboratory. In

total, each participant spent approximately 90 minutes in the laboratory, with ~50 minutes of this completing the main experiment.

3.2.5) Data Processing and Statistical Analysis

3.2.5.1) Electroencephalography Data Processing

The EEG processing for this study was conducted over 3 stages using the EEGLAB toolbox [261] for MATLAB [262] with the ERPLAB plugin [263]. The first processing stage was the basic processing, which began with converting the data to a format usable by EEGLAB from biowav to the standard EEGLAB '.set' format, and inserting electrode location data. The data was automatically referenced to the average of the physically linked-ears reference electrodes used during the EEG recording. A 0.5Hz Butterworth highpass filter was applied to counter slow frequency signal drift, and a 50Hz notch filter was applied to remove line noise. A temporary 40Hz lowpass filter was applied to aid with the identification of channels requiring interpolation, which was then executed on the non-40Hz filtered data.

The purpose of the second stage is to acquire independent component analysis (ICA) weights for eyeblink components, allowing eyeblinks to be selectively removed from the EEG data. However, isolating eyeblinks requires stricter filtering and artifact rejection criteria than are otherwise used in this processing procedure, which may unwantedly affect the data. Therefore, only the ICA weights are extracted from this stage for later application, and the datasets produced are not used in the final analysis. The dataset produced from stage 1 has a 40Hz filter applied to remove high-frequency noise. The data is then epoched from -200 to 1004ms (to account for a frame drop caused by rounding) based on the presentation onset of the question. The epochs are then manually reviewed, and any containing non-eyeblink related noise or perturbations are rejected, leaving otherwise clean data with eyeblink artifacts. The ICA process was then conducted on this data, which can selectively identify the eyeblinks with the EEG data and save the ICA weightings.

The third stage of the analysis produces the final ERP outputs. The dataset produced by stage 1 was epoched from -200 to 1004ms, and the ICA weights from stage 2 were applied. Components were identified using the ‘Label components’ function from the ‘Classify components using ICALabel’ menu in EEGLAB. Components that were labelled as originating from the eye were manually checked to see if they conform with typical eyeblink patterns. ICA components were then selectively removed individually and in each combination until a solution which removed the eyeblinks with minimal disruption to the EEG data was found. The data next underwent artifact rejection by visual inspection to remove remaining noise. The averaged ERPs for each condition were calculated individually for each participant, and averaged channels and conditions for statistical analysis were produced using the channel and bin operation functions, respectively. Every dataset was then combined into a grand average, where the mean amplitudes of the time windows for each targeted ERP component were extracted. The entire EEG/ERP analysis pipeline is shown in Figure 3.8. As no lowpass filter had been applied to this final dataset, a 30Hz filtered version of the computed averages and grand average was also produced for presentation purposes.

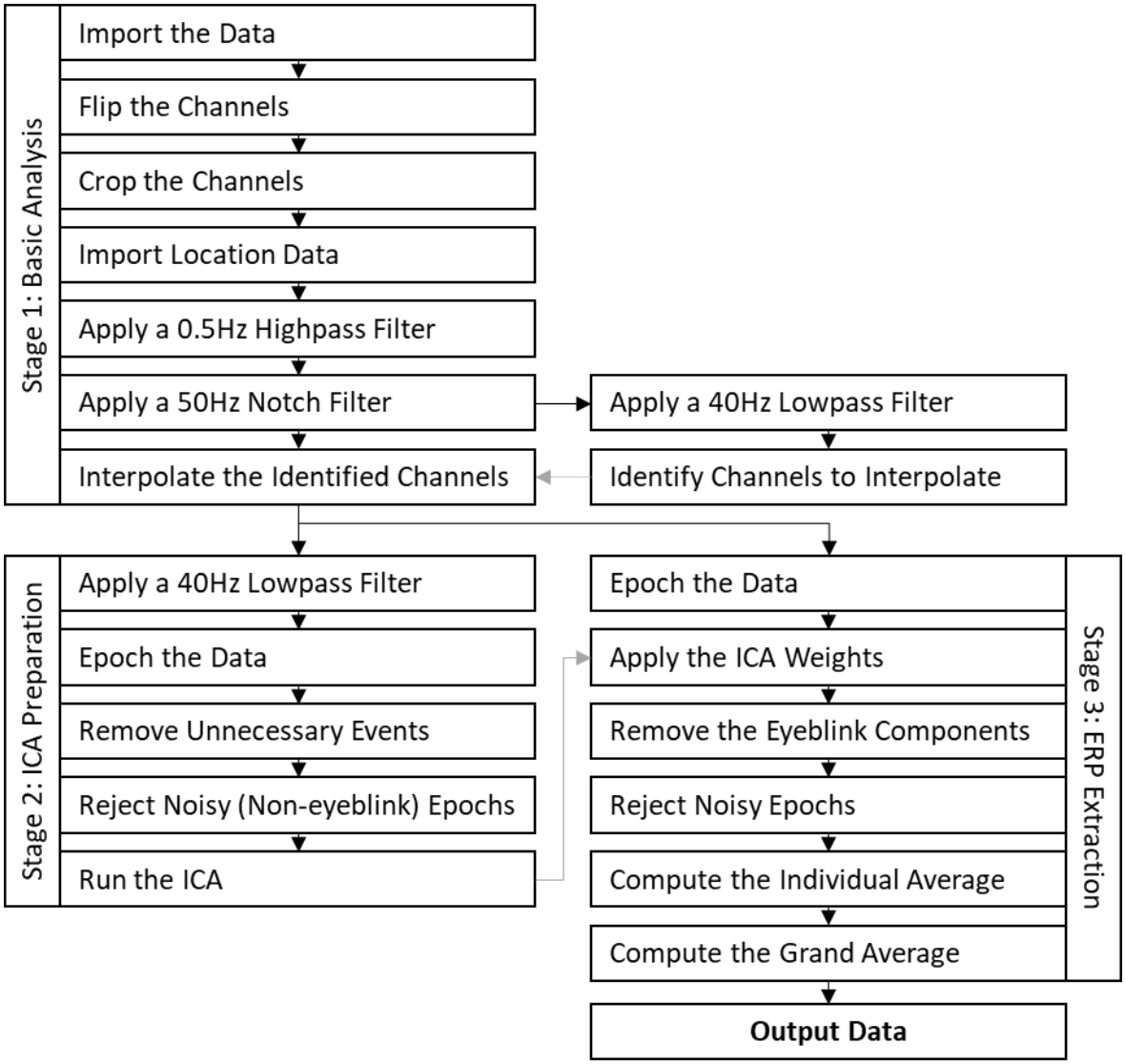


Figure 3.8: *EEG analysis chart showing all 3 stages of preprocessing procedure. Solid black lines show where the entire dataset is used in the next stage. The dotted grey line connecting the 'Identify Channels to Interpolate' to "Interpolate the Identified Channels" in stage 1 represents information from the former stage being used in the latter, but no actual data. The solid grey line collecting "Run the ICA" in stage 2 to "Apply the ICA Weights" in stage 3 represents part of the data from the former (the ICA weights calculated) being applied to the latter.*

3.2.5.2) Event Related Potential Time Window Selection

As this study is a novel investigation examining HMD-VR and mental arithmetic, the time windows for each component were visually inspected before statistical testing was conducted to ensure the responses were fully captured. The initial time windows used for the analysis of each targeted ERP component was based on published investigations to identify an appropriate range that peaks have been previously

reported (see Muluh [131] for a review), primarily Núñez-Peña et al. [221], Muluh [131], Jasinski & Coch [237], Moore et al. [241], and Dickson & Federmeier [223]. Specifically, the time windows used were between 100ms and 200ms for the N170, 250ms and 500ms for the P300, and 400ms and 800ms for the SWC. These time windows were considered the initial upper and lower bounds of reported components common to arithmetic tasks. These windows were applied to averaged ERPs within a presentation modality for each participant individually, and confirmed that the peak latency of the respective polarity extracted was observable within this range. At the final stage, the mean of the highest and lowest latency was identified, and a window length was selected to encompass all peaks within the range. When required, minor adjustments were made to round the window to the nearest 5ms.

The visual and auditory time windows are different due to different cognitive processes, but also the disparity between the onsets of the stimulus, caused by the 50ms of silence at the start of each audio files. The issue was corrected by shifting the time windows 115ms for the N170 component and 125ms for the P300 and SWC components within the auditory conditions. Moreover, the N170 time range had a fixed lower limit of 120ms due to a recording artifact associated with the triggers being present (Figure 3.9).

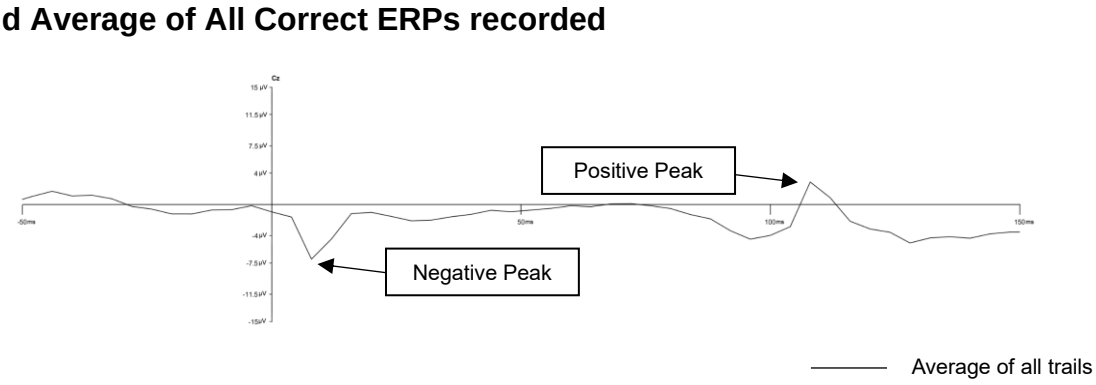


Figure 3.9: *Example of the trigger artifact: This ERP trace (-50 to 150ms) from the Fz electrode is extracted from the grand average of all correct trials regardless of condition or difficulty. This demonstrates the presence of a negative peak in the 10-20ms range and a positive peak within the 105-115ms (highlighted with arrows) range regardless of the stimuli or difficulties utilized. Because this artifact persists until the 115ms point, the lower limit of any time window used for analysis was selected as 120ms.*

The final time windows are as follows: The N170 time window is an 80ms time window centred at 160ms in the visual condition and 275ms in the auditory condition; the P300 time window is a 200ms window centred on 300ms in the visual condition, and 425ms in the auditory condition; the SWC is a 300ms window centred at 600ms in the visual condition, and 725ms in the auditory condition.

3.2.5.3) Statistical Analysis

Statistical analysis for behavioural and electrophysiological data was conducted in SPSS 28 [264] and Microsoft Excel [265]. For the behavioural results, the averaged response time of correctly answered questions, the average subjective difficulty rating, and the number of correct responses was calculated for each participant for each condition. The behavioural measures were individually analysed using a series of t-tests comparing between difficulties within the visual and auditory modalities. Incorrect responses were excluded from the response time analysis to prevent outliers resulting from immediate 'pass' responses (from pressing the enter button immediately without attempting to answer the question) from being included in the data. The responses to the post-experiment SSQ are grouped based on the subscales of nausea, oculomotor disturbance, disorientation and total score [160], which are analysed using a one-sample t-test against zero [266].

There was a low number of incorrect trials in each condition, with less than 10% (6) of responses being incorrect for each participant in the easy visual and easy auditory conditions, in all but 3 participants for the hard visual, and all but 5 participants for the hard auditory condition. Therefore, there was not enough results to warrant comparison within the incorrect responses nor against the correct responses [205,244]. These trials were therefore removed from the behavioural response time and ERP analyses.

Three types of ERP analysis are utilised: mean amplitude, peak amplitude, and peak latency. Each ERP analysis was conducted on each of the targeted time windows individually for correct responses only, but separately for each modality. Six ANOVA

were conducted per analysis method, totalling 18 across all ERP analyses. Each ANOVA consisted of a 3-factor repeated measures design, using a 2x3x3 (difficulty [easy, hard] x electrode location [frontal, central, parietal] x hemisphere [left, middle, right]) design. Mauchly's test of sphericity was employed to identify if any main effects and interactions violated the sphericity assumption. When sphericity is violated, the degrees of freedom of the effect is adjusted using either the Greenhouse-Geisser correction when the Greenhouse-Geisser Epsilon is under 0.75, and the Huynh-Feldt correction when the Greenhouse-Geisser Epsilon is over 0.75. Post-hoc analysis was conducted using pairwise comparisons, corrected for multiple comparisons using the Bonferroni adjustment. The Bonferroni correction was used to control for type 1 error when calculating the post hoc comparisons of the main effects and interactions. This was selected based on the recommendation of Field [267], particularly because of the repeated measures design, low number of comparisons within each factor, and the fact that sphericity could often not be assumed. As each behavioural results comparison only had two levels, paired t-tests were used.

3.3) Results

3.3.1) Behavioural

Paired samples T-tests were used for the behavioural comparisons between easy and hard conditions for each display type as each only contained 2 levels. One-sample T-tests were used for the SSQ scores.

3.3.1.1) Number of Correct Reponses

The behavioural results of the mean number of correct responses (Figure 3.10) find that within the visual trials, there were significantly more correct responses for the easy-visual (Mean ± SEM (M) = 59.48 ± 0.16, SD = 0.75, range = 58-60) than the hard-visual (M = 56.48 ± 0.79, SD = 3.64, range = 44-60) trials ($t(20)=4.31$, $p \leq 0.001$, $d=0.94$). Within the auditory trials, there were significantly more correct responses in the easy-auditory (M = 59.33 ± 0.2, SD = 0.91, range = 57-60) over the hard-auditory (M = 55.14 ± 0.75, SD = 3.42, range = 47-60) trials ($t(20)=6.15$, $p \leq 0.001$, $d=1.34$).

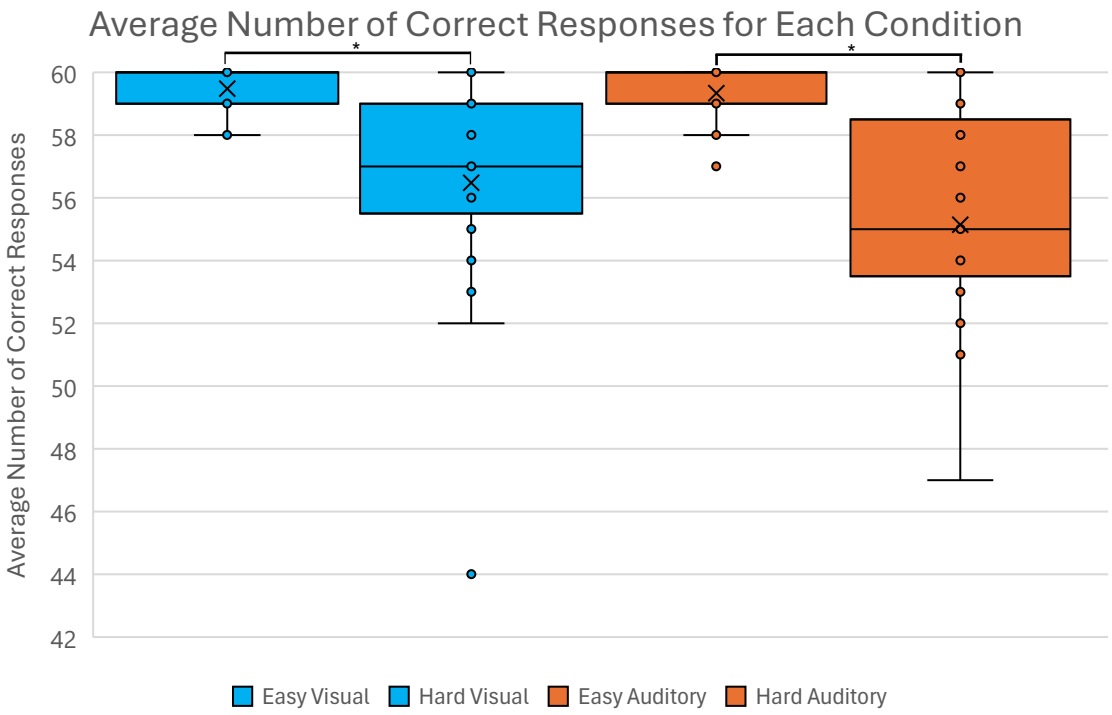


Figure 3.10: *A box plot showing the average number of correct responses per condition for all participants. The maximum number of correct responses was 60. n=21.*
**Significant differences between conditions, p<0.05.*

3.3.1.2) Average Response Time

The behavioural measure of the response time for correct answers (Figure 3.11) finds that within the visual trials, response times for the easy-visual (M = 2.61 ± 0.13, SD = 0.61, range = 1.85-3.97) trials were significantly faster than in the hard-visual (M = 4.57 ± 0.36, SD = 1.66, range = 2.99-8.88) ($t(20)=-7.42$, $p\leq0.001$, $d=-1.62$). Within the auditory trials, it was found that the easy-auditory (M = 2.69 ± 0.14, SD = 0.66, range = 1.78-4.18) was significantly faster than the hard-auditory (M = 7.68 ± 0.71, SD = 3.26, range = 4.15-15.7) trials ($t(20)=-8.18$, $p\leq0.001$, $d=-1.79$).

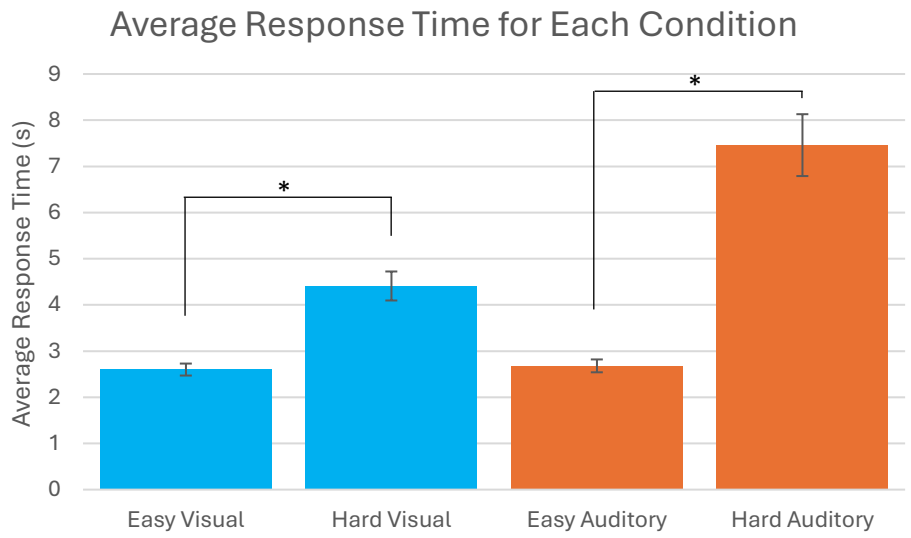


Figure 3.11: *A bar chart showing the average response time to the answer input in seconds for all conditions. n=21.*
**Significant differences between conditions, p<0.05.*

3.3.1.3) Subjective Difficulty Rating

The behavioural measure of subjective difficulty rating (Figure 3.12) found that, within the visual trials, it was found that the easy-visual (M = 27.82 ± 5.19, SD = 23.78, range = 3.93-103.93) trials were rated at a significantly lower difficulty than the hard-visual (M = 195.02 ± 19.75, SD = 90.5, range = 40.12-343.2) trials (t(20)=-9.06, p≤0.001, d=-1.98). Within the auditory trials, it was found that the easy-auditory (M = 29.52 ± 6.32, SD = 28.98, range = 3.88-123.18) trials were rated significantly lower difficulty than the hard-auditory (M = 239.27 ± 22.26, SD = 102.02, range = 71.58-417.48) trials (t(20)=-9.91, p≤0.001, d=-2.16).

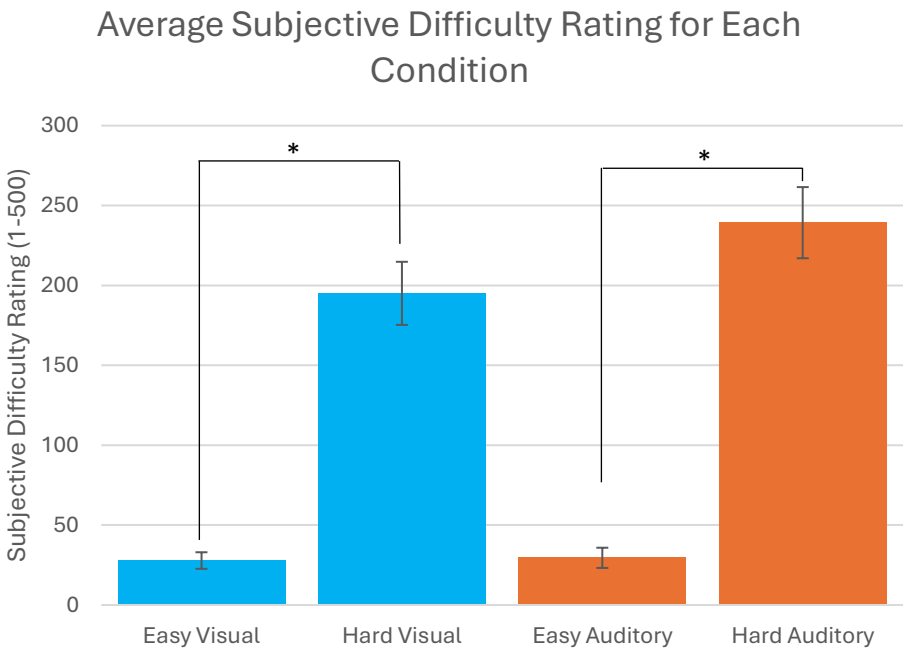


Figure 3.12: *A bar chart showing the average subjective difficulty rating for each condition on a scale of 1-500. n=21.*
**Significant differences between conditions, p<0.05.*

3.3.1.4) Simulator Sickness Questionnaire

One-sample t-tests found significant increases for all subscales of the post-experiment SSQ compared to no change: Nausea (M=39.52 ± 4.83; t(20)=8.07, p≤0.001), Oculomotor Disturbance (M=59.56 ± 6.56; t(20)=8.91, p≤0.001), Disorientation (M=45.07 ± 6.64; t(20)=6.58, p≤0.001), Total Score (M=56.99 ± 6.33; t(20)=8.81, p≤0.001).

3.3.2) EEG Event-Related Potentials

Individual repeated measures ANOVA (2x3x3 design (difficulty [easy, hard] x electrode location [frontal, central, parietal] x hemisphere [left, midline, right]) were conducted for each combination of ERP type (mean amplitude, peak amplitude and peak latency), modality (visual and auditory) and ERP component (N170, P300, SWC). The targeted components are analysed based on the modality-specific time windows identified for each component.

3.3.2.1) Visual N170 Component

Separate traces were produced of the average visual ERPs for 9 electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4) for both difficulty conditions (easy and hard) separately (Figure 3.13). The statistical analyses for the mean amplitude, peak amplitude and peak latency component of the visual N170 component are summarised in Table 3.1. In the visual N170 mean amplitude analysis, Mauchly's Test of Sphericity is violated for the interaction between difficulty x electrode ($\chi^2(2)=13.153$, $p=0.001$, $\varepsilon=0.667$) only, and is adjusted using the Greenhouse-Geisser correction. The main effect of electrode is significant, with post hoc comparisons demonstrating that central ($M=-2.84 \pm 0.62$) and parietal ($M=-3.68 \pm 0.58$) had larger mean amplitudes than the frontal ($M=-1.77 \pm 0.51$) electrodes ($t(20)=1.07$, $p=0.008$, $d=0.74$; $t(20)=1.91$, $p=0.004$, $d=0.80$). No interactions are significant.

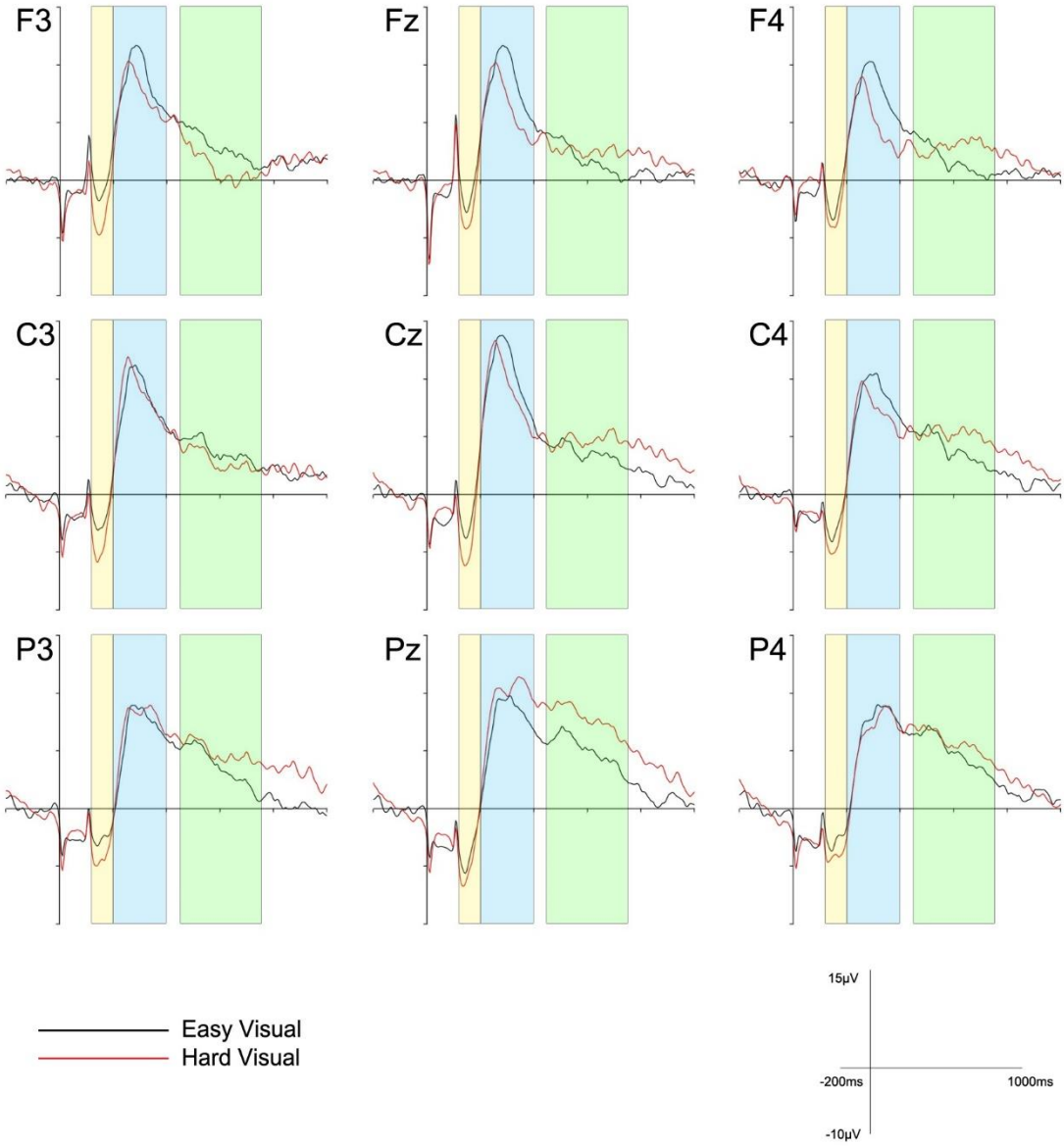


Figure 3.13: *The grand average of all 21 participants for the Easy-Visual and Hard-Visual conditions for the 9 electrodes examined (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4). The respective time window for each component is overlaid: The yellow time window is the N170 time range (120-200ms), the blue time window is the P300 (200-400ms), and the green is the SWC time window (450-750ms). A 30Hz lowpass filter has been applied for clarity. n=21.*

Table 3.1: *The ERP ANOVA results for the mean amplitude, peak amplitude and peak latency N170 responses in the visual condition. n=21.*

	Mean Amplitude				Peak Amplitude				Peak Latency			
Within Subject Effect	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**
Difficulty	4.05	1, 20	0.06	0.17	7.73	1, 20	**0.012***	0.28	1.07	1, 20	0.31	0.05
Electrode	9.22	2, 40	**0.001***	0.32	4.95	2, 40	**0.012***	0.2	0.6	1.43, 28.62	0.5	0.03
Hemisphere	0.23	2, 40	0.8	0.01	0.4	2, 40	0.67	0.02	2.37	1.61, 32.16	0.12	0.11
Difficulty x Electrode	0.3	1.33, 26.68	0.66	0.02	0.04	1.39, 27.87	0.91	<0.01	0.12	1.36, 27.22	0.81	0.01
Difficulty x Hemisphere	1.29	2, 40	0.29	0.06	3.21	2, 40	0.05	0.14	3.16	2, 40	0.05	0.14
Electrode x Hemisphere	1.18	4, 80	0.33	0.06	2.46	4, 80	0.05	0.11	1.23	4, 80	0.3	0.06
Difficulty x Electrode x Hemisphere	0.91	4, 80	0.46	0.04	1.61	4, 80	0.18	0.07	0.6	2.51, 50.25	0.59	0.03

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

In the visual N170 peak amplitude analysis, Mauchly's Test of Sphericity is violated for the interaction between difficulty x electrode ($\chi^2(2)=10.86$, p=0.004, ε=0.7) only, and is adjusted using the Greenhouse-Geisser correction. The main effects of difficulty and electrode are significant. Post hoc comparisons found that between difficulties, the hard peak amplitude (M=-8.18 ± 0.87) is larger than the easy peak amplitude (M=-5.94 ± 0.63) condition (t(20)=2.25, p=0.012, d=0.61). Between electrodes, the parietal peak amplitude (M=-7.84 ± 0.76) is larger than the frontal peak amplitude (M=-6.27 ± 0.64) (t(20)=1.57, p=0.04, d=0.59), but there are no differences from the central electrode.

In the visual N170 peak latency analysis, Mauchly's Test of Sphericity is violated for the main effect of electrode ($\chi^2(2)=9.63$, p=0.008, ε=0.72) and the interactions of difficulty x electrode ($\chi^2(2)=12.05$, p=0.002, ε=0.68) & difficulty x electrode x hemisphere ($\chi^2(9)=20.20$, p=0.017, ε=0.63), which are adjusted using the

Greenhouse-Geisser correction. Moreover, Mauchly's Test of Sphericity is violated for the main effect of hemisphere ($\chi^2(2)=7.43$, $p=0.024$, $\varepsilon=0.76$), which is adjusted using the Hyunh-Feldt correction. However, no significant main effects or interactions are found.

3.3.2.2) Visual P300 Component

The statistical analyses for the mean amplitude, peak amplitude and peak latency component of the visual P300 component are summarised in table 3.2. In the visual P300 mean amplitude analysis, Mauchly's Test of Sphericity is violated for the main effects of electrode ($\chi^2(2)=8.65$, $p=0.013$, $\varepsilon=0.73$) & hemisphere ($\chi^2(2)=8.62$, $p=0.013$, $\varepsilon=0.73$) and interactions electrode x hemisphere ($\chi^2(9)=21.22$, $p=0.012$, $\varepsilon=0.71$) & difficulty x electrode x hemisphere ($\chi^2(9)=22.23$, $p=0.008$, $\varepsilon=0.75$). In all violations the Greenhouse-Giesser correction is used. The main effect of hemisphere is significant, with post hoc comparisons finding the midline mean amplitude ($M=8.40 \pm 0.83$) is larger than the right mean amplitude ($M=6.67 \pm 0.76$) ($t(20)=1.74$, $p\leq0.001$, $d=1.04$).

The interaction between difficulty x electrode is also significant, with post hoc comparisons finding that within electrode location and between difficulties, frontal-easy ($M=8.01 \pm 0.73$) mean amplitude is larger than the frontal-hard mean amplitude ($M=6.24 \pm 0.97$) ($t(20)=1.77$, $p=0.007$, $d=0.65$). When comparing within difficulties and between electrode locations, there are differences between the easy-central ($M=8.67 \pm 0.74$) & easy-parietal ($M=6.77 \pm 0.94$) mean amplitudes ($t(20)=1.9$, $p=0.032$, $d=0.62$) and hard-frontal ($M=6.24 \pm 0.97$) & hard-central ($M=8.05 \pm 1.14$) mean amplitudes ($t(20)=-1.81$, $p=0.028$, $d=-0.63$), with the central electrode mean amplitude being larger in both cases.

Table 3.2: *The ERP ANOVA results for the mean amplitude, peak amplitude and peak latency P300 responses in the visual condition. n=21.*

	Mean Amplitude				Peak Amplitude				Peak Latency			
Within Subject Effect	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**
Difficulty	0.66	1, 20	0.43	0.03	0.48	1, 20	0.5	0.02	3.42	1, 20	0.08	0.15
Electrode	2.41	1.47, 29.29	0.12	0.11	2.57	2, 40	0.09	0.11	10.73	2, 40	**<0.001***	0.35
Hemisphere	5.81	1.47, 29.31	**0.013***	0.23	6.6	2, 40	**0.003***	0.25	6.17	2, 40	**0.005***	0.24
Difficulty x Electrode	7.22	2, 40	**0.002***	0.27	1.38	2, 40	0.26	<0.01	2.47	2, 40	0.1	0.11
Difficulty x Hemisphere	1.77	2, 40	0.18	0.08	0.39	2, 40	0.68	0.02	0.47	2, 40	0.63	0.02
Electrode x Hemisphere	1.77	2.84, 56.89	0.17	0.08	2.59	4, 80	**0.043***	0.12	1.74	2.48, 49.62	0.18	0.08
Difficulty x Electrode x Hemisphere	2.1	2.99, 59.69	0.11	0.1	2.01	4, 80	0.1	0.09	0.93	4, 80	0.45	0.04

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

In the visual P300 peak amplitude analysis, Mauchly's Test of Sphericity is not violated on any main effect or interaction. The main effect of hemisphere is significant, with post hoc comparisons finding the midline hemisphere peak amplitude (M=15.79 ± 1.19) is larger than the right hemisphere peak amplitude (M=13.7 ± 1.02) (t(20)=2.09, p=0.001, d=1). The interaction between electrode x hemisphere is also significant, with post hoc comparisons within hemisphere and between electrodes finding a significantly larger peak amplitude in the midline-central (M=17.44 ± 1.44) compared to the midline-frontal (M=14.34 ± 1.27) (t(20)=-3.10, p=0.001, d=-0.99). Moreover, the post hoc comparisons within electrodes and between hemispheres find the central-midline peak amplitude (M=17.44 ± 1.44) is larger than the central-right peak amplitude (M=14.29 ± 1.2) (t(20)=3.15, p≤0.001, and the parietal-midline peak amplitude (M=15.58 ± 1.14) is larger than parietal-right peak amplitude (M=13.28 ± 1.18) (t(20)=2.29, p=0.001, d=0.97).

In the visual P300 peak latency analysis, Mauchly's Test of Sphericity is violated for the interaction between electrode x hemisphere ($\chi^2(9)=22.22$, $p=0.008$, $\varepsilon=0.62$), and is adjusted using the Greenhouse-Geisser correction. The main effects of electrode and hemisphere are significant. Post hoc comparisons between electrodes found the parietal peak (M=311.91 ± 8.11) occurred later than both the frontal (M=282.25 ± 5.51) (t(20)=-29.65, p=0.003, d=-0.83) and central (M=291.97 ± 7.13) (t(20)=-19.94, p=0.007, d=-0.77) peaks. Between hemispheres, it is found the peak in the right hemisphere (M=302.19 ± 5.78) is later than the midline (M=291.08 ± 6.15) (t(20)=-11.11, p=0.009, d=-0.74).

3.3.2.3) Visual Slow Wave Component

The statistical analyses for the mean amplitude, peak amplitude and peak latency component of the visual SWC are summarised in table 3.3. For the visual SWC mean amplitude analysis, Mauchly's Test of Sphericity is violated for the main effects of electrode ($\chi^2(2)=9.30$, $p=0.01$, $\varepsilon=0.72$) & hemisphere ($\chi^2(2)=7.90$, $p=0.019$, $\varepsilon=0.75$), and the interaction between difficulty x electrode ($\chi^2(2)=13.94$, $p=0.001$, $\varepsilon=0.66$). Each violation is adjusted using the Greenhouse-Geisser correction. The main effect of electrode is significant, with post hoc tests finding that the frontal mean amplitude (M=2.15 ± 0.65) is smaller than both the central (M=3.99 ± 0.74) (t(20)=-1.84, p=0.012, d=-0.71) and parietal (M=5.35 ± 0.87) (t(20)=-3.2, p=0.005, d=-0.79) mean amplitudes.

Table 3.3: *The ERP ANOVA results for the mean amplitude, peak amplitude and peak latency slow wave component responses in the visual condition. n=21.*

	Mean Amplitude				Peak Amplitude				Peak Latency			
Within Subject Effect	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**
Difficulty	0.75	1, 20	0.4	0.04	1.12	1, 20	0.3	0.05	2.47	1, 20	0.13	0.11
Electrode	10.76	1.44, 28.84	**0.001***	0.35	5.99	1.27, 25.34	**0.016***	0.23	1.39	2, 40	0.26	0.07
Hemisphere	2.83	1.49, 29.84	0.09	0.12	0.37	1.29, 25.75	0.6	0.02	1.11	1.41, 28.18	0.32	0.05
Difficulty x Electrode	5.01	1.32, 26.32	**0.025***	0.2	2.55	1.29, 25.83	0.12	<0.01	3.28	2, 40	**0.048***	0.14
Difficulty x Hemisphere	4.71	2, 40	**0.015***	0.19	2.02	1.49, 29.82	0.16	0.09	0.02	1.64, 32.85	0.97	<0.01
Electrode x Hemisphere	2.12	4, 80	0.09	0.1	2.11	2.51, 50.15	0.12	0.1	3.45	4, 80	**0.012***	0.15
Difficulty x Electrode x Hemisphere	2.36	4, 80	0.06	0.11	1.69	4, 80	0.16	0.08	1.99	4, 80	0.1	0.09

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

The visual SWC mean amplitude analysis also found a significant interaction between difficulty x electrode. Post hoc comparisons within difficulties and between electrodes found that, easy-frontal (M=2.14 ± 0.73) mean amplitudes are smaller than both the easy-central (M=3.73 ± 0.82) (t(20)=-1.59, p=0.033, d=-0.61) and easy-parietal (M=4.56 ± 0.97) (t(20)=-2.41, p=0.047, d=-0.58) mean amplitudes. Moreover, the hard-frontal mean amplitude (M=2.16 ± 0.83) is smaller than both the hard-central (M=4.25 ± 0.87) (t(20)=-2.1, p=0.008, d=-0.75) and hard-parietal (M=6.14 ± 1) (t(20)=-3.98, p=0.002, d=-0.90) mean amplitudes, which in turn differed with the hard-parietal being larger than the hard-central mean amplitude (t(20)=-1.88, p=0.012, d=-0.71). Furthermore, the interaction between difficulty x hemisphere is significant, with post hoc comparisons revealing that the hard-left mean amplitude (M=2.97 ± 0.84) is smaller than the hard-midline mean amplitude (M=4.97 ± 0.91) (t(20)=-1.99, p=0.006, d=-0.78).

For the visual SWC peak amplitude analysis, Mauchly's Test of Sphericity is violated for the main effects of electrode ($\chi^2(2)=16.41$, $p<.01$, $\varepsilon=0.63$) & hemisphere ($\chi^2(2)=15.33$, $p<.01$, $\varepsilon=0.64$) and the interactions between difficulty x electrode ($\chi^2(2)=15.12$, $p=0.001$, $\varepsilon=0.65$), difficulty x hemisphere ($\chi^2(2)=7.94$, $p=0.019$, $\varepsilon=0.75$) & electrode x hemisphere ($\chi^2(9)=21.94$, $p=0.009$, $\varepsilon=0.63$). All violations are adjusted using the Greenhouse-Giesser correction. The main effect of electrode is significant, with the post hoc comparisons finding the central peak amplitude (M=10.77 ± 0.97) is larger than the frontal peak amplitude (M=9.21 ± 0.96) (t(20)=-1.56, p=0.041, d=-0.59).

For the visual SWC peak latency analysis, Mauchly's Test of Sphericity is violated for the main effect of hemisphere ($\chi^2(2)=10.33$, $p=0.006$, $\varepsilon=0.71$) and the interaction of difficulty x hemisphere ($\chi^2(2)=6.77$, $p=0.034$, $\varepsilon=0.77$), which are adjusted with the Greenhouse-Giesser and Huynh-Feldt corrections respectively. The interactions between difficulty x electrode and electrode x hemisphere are significant. Post hoc comparisons of the difficulty x electrode interaction find that, within difficulties and between electrodes, the hard-central peak (M=592.83 ± 16.34) is significantly later than the hard-parietal peak (M=555.05 ± 14.96) (t(20)=37.78, p=0.032, d=0.61). Post hoc comparisons of the electrode x hemisphere interaction find that, within hemispheres and between electrodes the midline-central peak (M=601.24 ± 14.32) is significantly later than the midline-parietal peak (M=556 ± 13.88) (t(20)=45.24, p=0.014, d=0.7). Comparisons within electrodes and between hemispheres find the central-midline peak (M=601.24 ± 14.32)) is later than the central-left (M=566.76 ± 16.04) (t(20)=-34.48, p=0.046, d=-0.58) and central-right (M=560.86 ± 13.09) (t(20)=40.38, p=0.001, d=0.95) peaks.

3.3.2.4) Auditory N170 Component

Separate traces were produced of the average auditory ERPs for 9 electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4) for both difficulty conditions (easy and hard) separately (Figures 3.14). The statistical analyses for the mean amplitude, peak amplitude and peak latency component of the auditory N170 component are summarised in table 3.4. For the auditory N170 mean amplitude analysis, Mauchly's

Test of Sphericity is violated by the main effect of electrode ($\chi^2(2)=6.14$, $p=0.046$, $\varepsilon=0.78$) and the interaction of difficulty x electrode ($\chi^2(2)=6.19$, $p=0.045$, $\varepsilon=0.78$), which are adjusted using the Huynh-Feldt correction. Moreover, Mauchly's Test of Sphericity is also violated by the interactions of difficulty x hemisphere ($\chi^2(2)=14.82$, $p=0.001$, $\varepsilon=0.65$) & difficulty x electrode x hemisphere ($\chi^2(9)=31.58$, $p<.01$, $\varepsilon=0.54$), which are adjusted using the Greenhouse-Giesser correction. The main effect of electrode is significant, with post hoc comparisons finding that the frontal mean amplitude ($M=-2.18 \pm 0.99$) is smaller than the central ($M=-3.81 \pm 0.91$) ($t(20)=1.63$, $p=0.002$, $d=0.86$) and parietal ($M=-4.63 \pm 0.76$) ($t(20)=2.45$, $p=0.001$, $d=0.91$) mean amplitudes.

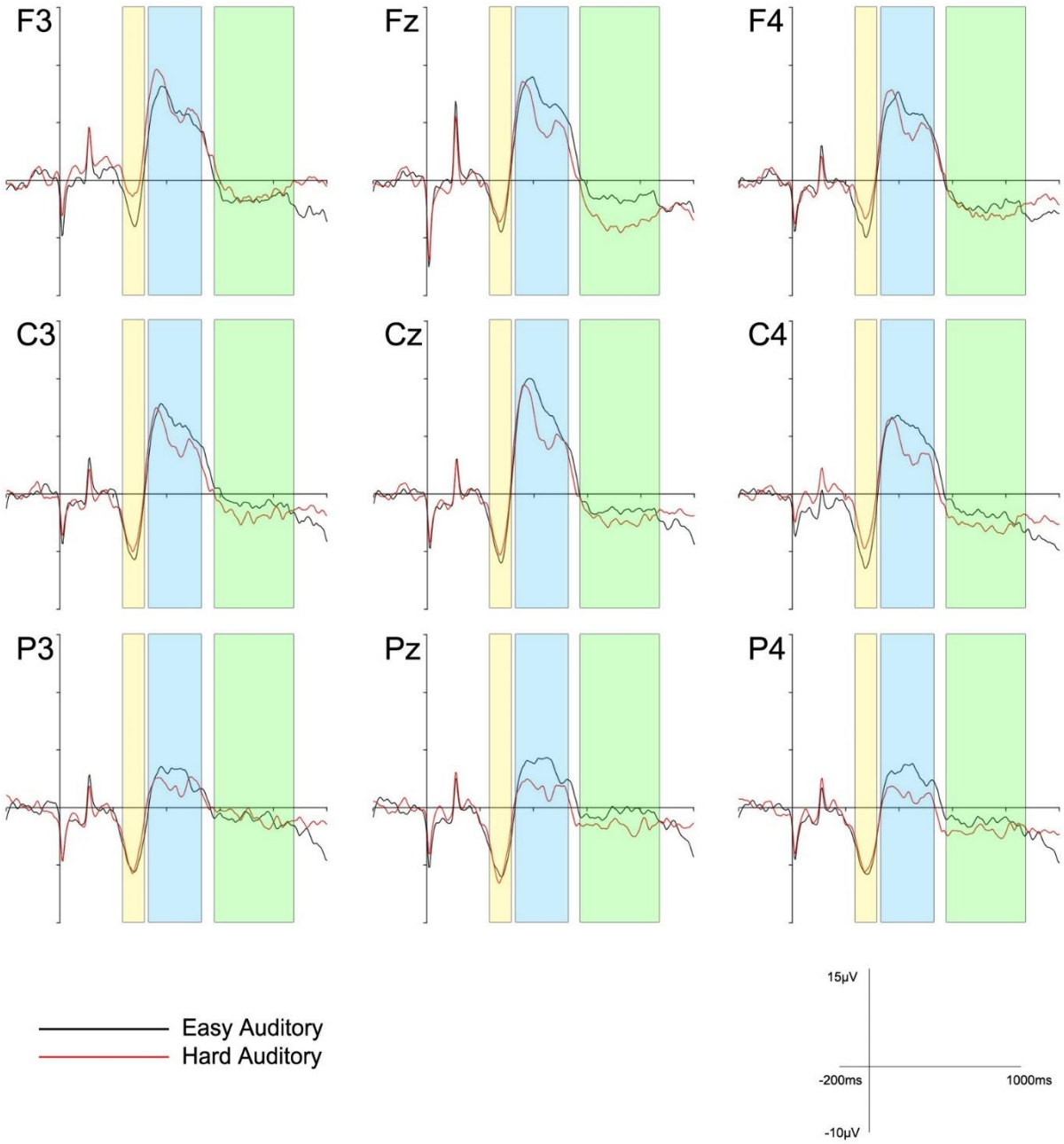


Figure 3.14: *The grand average of all 21 participants for the Easy-Auditory and Hard-Auditory conditions for the 9 electrodes examined (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4). The respective time window for each component is overlaid: The yellow time window is the N170 time range (235-315ms), the blue time window is the P300 (325-525ms), and the green is the SWC time window (575-875ms). A 30Hz lowpass filter has been applied for clarity.*

Table 3.4: *The ERP ANOVA results for the mean amplitude, peak amplitude and peak latency N170 responses in the auditory condition. n=21.*

	Mean Amplitude				Peak Amplitude				Peak Latency			
Within Subject Effect	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**
Difficulty	1.36	1, 20	0.26	0.06	1.8	1, 20	0.2	0.08	0.1	1, 20	0.76	0.01
Electrode	13.73	1.68, 33.54	**<0.001***	0.41	7.21	1.37, 27.43	**0.007***	0.27	0.86	1.45, 28.9	0.4	0.04
Hemisphere	2.11	2, 40	0.14	0.1	1.85	2, 40	0.17	0.09	1.18	2, 40	0.32	0.06
Difficulty x Electrode	1.43	1.67, 33.49	0.25	0.07	1.04	1.66, 33.15	0.35	<0.01	0.36	2, 40	0.7	0.02
Difficulty x Hemisphere	0.63	1.3, 25.95	0.48	0.03	0.38	1.39, 27.85	0.61	0.02	0.23	2, 40	0.8	0.01
Electrode x Hemisphere	1.11	4, 80	0.36	0.05	2.23	4, 80	0.07	0.1	0.96	4, 80	0.44	0.05
Difficulty x Electrode x Hemisphere	0.76	2.16, 43.2	0.48	0.04	1.18	1.98, 39.67	0.32	0.06	0.39	4, 80	0.82	0.02

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

For the auditory N170 peak amplitude analysis, Mauchly's Test of Sphericity is violated by the main effect of electrode ($\chi^2(2)=11.64$, p=0.003, ε=0.69) and the interactions of difficulty x hemisphere ($\chi^2(2)=10.89$, p=0.004, ε=0.7) & difficulty x electrode x hemisphere ($\chi^2(9)=34.57$, p≤0.001, ε=0.5), which are adjusted using the Greenhouse-Giesser correction. Moreover, Mauchly's Test of Sphericity is also violated by the interaction of difficulty x electrode ($\chi^2(2)=6.49$, p=0.039, ε=0.78), which is adjusted using the Huynh-Feldt correction. The main effect of electrode is significant, with post hoc comparisons finding a larger peak amplitude in the central electrode (M=-8.82 ± 1.14) than the frontal electrode (M=-6.91 ± 1.26) (t(20)=1.91, p=0.001, d=0.98).

For the auditory N170 peak latency analysis, Mauchly's Test of Sphericity is violated by the main effect of electrode ($\chi^2(2)=9.21$, p=0.01, ε=0.72), and is adjusted using the Greenhouse-Giesser correction. However, no significant main effects or interactions are found.

3.3.2.5) Auditory P300 Component

The statistical analyses for the mean amplitude, peak amplitude and peak latency component of the auditory P300 component are summarised in table 3.5. For the auditory P300 mean amplitude analysis, Mauchly's Test of Sphericity is violated for the main effect of electrode ($\chi^2(2)=6.87$, $p=0.032$, $\varepsilon=0.77$) and the interaction between difficulty x electrode ($\chi^2(2)=6.85$, $p=0.033$, $\varepsilon=0.77$), which are adjusted using the Huynh-Feldt correction. Moreover, Mauchly's Test of Sphericity is violated for the interaction of difficulty x hemisphere ($\chi^2(2)=9.15$, $p=0.01$, $\varepsilon=0.72$), which is adjusted using the Greenhouse-Giesser correction. The main effect of electrode is significant, with post hoc comparisons finding the parietal mean amplitude (M=2.12 ± 0.70) is smaller than the frontal (M=6.03 ± 0.84) ($t(20)=3.91$, $p\leq0.001$, $d=1.44$) and central (M=5.57 ± 0.80) ($t(20)=3.45$, $p\leq0.001$, $d=1.99$) mean amplitudes.

Table 3.5: *The ERP ANOVA results for the mean amplitude, peak amplitude and peak latency P300 responses in the auditory condition. n=21.*

	Mean Amplitude				**Peak Amplitude**				**Peak Latency**			
Within Subject Effect	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**
Difficulty	1.36	1, 20	0.26	0.06	0.17	1, 20	0.68	0.01	0.05	1, 20	0.82	<0.01
Electrode	39.85	1.64, 32.73	**<0.001***	0.67	38.17	2, 40	**<0.001***	0.66	8.28	1.49, 29.7	**0.003***	0.29
Hemisphere	2.75	2, 40	0.08	0.12	4.82	2, 40	**0.013***	0.19	0.92	2, 40	0.41	0.04
Difficulty x Electrode	1.17	1.64, 32.76	0.31	0.06	1.15	1.67, 33.36	0.32	<0.01	1.08	2, 40	0.35	0.05
Difficulty x Hemisphere	1.89	1.45, 28.94	0.18	0.09	2.98	1.31, 26.12	0.09	0.13	0.07	1.65, 32.93	0.9	<0.01
Electrode x Hemisphere	1.57	4, 80	0.19	0.07	3.4	4, 80	**0.013***	0.15	1.36	4, 80	0.26	0.06
Difficulty x Electrode x Hemisphere	0.67	4, 80	0.62	0.03	0.96	4, 80	0.43	0.05	0.55	4, 80	0.7	0.03

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, $p<0.05$.

For the auditory P300 peak amplitude analysis, Mauchly's Test of Sphericity is violated for the interactions of difficulty x electrode ($\chi^2(2)=6.3$, $p=0.043$, $\varepsilon=0.78$), which is adjusted using the Huynh-Feldt correction; and difficulty x hemisphere ($\chi^2(2)=14.41$, $p=0.001$, $\varepsilon=0.653$) which is adjusted using the Greenhouse-Giesser correction. The main effects of electrode and hemisphere are significant. Post hoc comparisons find that between electrodes, the parietal peak amplitude (M=8.24 ± 0.79) is smaller than both the frontal (M=12.61 ± 1.15) (t(20)=4.37, p≤0.001, d=1.4) and central (M=12.5 ± 1.12) (t(20)=4.26, p≤0.001, d=2.08) electrode peak amplitudes. Between hemispheres, the midline peak amplitude (M=11.66 ± 1.12) is larger than the right peak amplitude (M=10.19 ± 1.01) (t(20)=1.47, p=0.02, d=0.66).

The auditory P300 peak amplitude analysis also found a significant interaction between electrode x hemisphere, with post hoc comparisons finding that within electrodes and between hemispheres, the central-midline peak amplitude (M=13.95 ± 1.36) is larger than the central-right peak amplitude (M=11.09 ± 1.16) (t(20)=2.86, p=0.007, d=0.76). Within hemispheres and between electrodes, it is found that the left-parietal peak amplitude (M=8.82 ± 0.82) is smaller than the left-frontal (M=13.21 ± 1.09) (t(20)=4.4, p≤0.001, d=1.28) and left-central (M=12.46 ± 1.1) (t(20)=3.64, p≤0.001, d=1.22) peak amplitudes; the midline-parietal peak amplitude (M=8.47 ± 0.92) is smaller than the midline-frontal (M=12.56 ± 1.31) (t(20)=4.1, p≤0.001, d=1.02) and midline-central (M=13.95 ± 1.36) (t(20)=5.48, p≤0.001, d=1.80) peak amplitudes; and that the right-parietal (M=7.442 ± 0.865) is smaller than the right-frontal (M=12.05 ± 1.18) (t(20)=4.61, p≤0.001, d=1.40) and right-central (M=11.09 ± 1.16) (t(20)=3.65, p≤0.001, d=1.39) peak amplitudes.

For the auditory P300 peak latency analysis, Mauchly's Test of Sphericity is violated for the main effect of electrode ($\chi^2(2)=8.1$, $p=0.017$, $\varepsilon=0.74$) and the interaction difficulty x hemisphere ($\chi^2(2)=6.68$, $p=0.035$, $\varepsilon=0.77$), which are adjusted with the Greenhouse-Geisser correction and Huynh-Feldt correction respectively. The main effect of electrode is significant, with post hoc comparisons showing the parietal peak (M=424.64 ± 10.32) is later than the frontal (M=399.14 ± 9.79) (t(20)=-25.49,

p=0.016, d=-0.68) and central (M=406.06 ± 10.69) (t(20)=-18.57, p=0.009, d=-0.74) peaks.

3.3.2.6) Auditory Slow Wave Component

The statistical analyses for the mean amplitude, peak amplitude and peak latency component of the auditory SWC are summarised in table 3.6. For the auditory SWC mean amplitude analysis, Mauchly's Test of Sphericity is violated for the main effect of electrode ($\chi^2(2)=9.937$, p=0.007, ε=0.711) and the interaction between difficulty x electrode ($\chi^2(2)=10.975$, p=0.004, ε=0.695), which were both adjusted with the Greenhouse-Giesser correction. However, no significant main effects or interactions are found.

Table 3.6: *The ERP ANOVA results for the mean amplitude, peak amplitude and peak latency slow wave component responses in the auditory condition. n=21.*

	Mean Amplitude				Peak Amplitude				Peak Latency			
Within Subject Effect	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**	**F**	**df**	**p**	**η_p^2**
Difficulty	1.27	1, 20	0.27	0.06	0.38	1, 20	0.55	0.02	0.56	1, 20	0.46	0.03
Electrode	1.41	1.42, 28.42	0.26	0.07	1.58	1.64, 32.83	0.22	0.07	0.04	2, 40	0.97	<0.01
Hemisphere	2.53	2, 40	0.09	0.11	2.46	2, 40	0.1	0.11	3.36	2, 40	**0.045***	0.14
Difficulty x Electrode	0.07	1.39, 27.8	0.87	<0.01	0.58	2, 40	0.56	<0.01	1.7	2, 40	0.2	0.08
Difficulty x Hemisphere	1.87	2, 40	0.17	0.09	1.5	2, 40	0.24	0.07	1.44	2, 40	0.25	0.07
Electrode x Hemisphere	0.54	4, 80	0.71	0.03	0.35	4, 80	0.85	0.02	1.29	4, 80	0.28	0.06
Difficulty x Electrode x Hemisphere	1.18	4, 80	0.32	0.06	0.72	4, 80	0.58	0.04	1.6	4, 80	0.18	0.07

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

For the auditory SWC peak amplitude analysis, Mauchly's Test of Sphericity is violated for the main effect of electrode ($\chi^2(2)=6.78$, $p=0.034$, $\varepsilon=0.77$), and is adjusted using the Huynh-Feldt correction. However, no significant main effects or interactions are found.

For the auditory SWC peak latency analysis, Mauchly's Test of Sphericity is not violated on any main effect or interaction. The main effect of hemisphere is significant, however post hoc tests find no significant differences between any comparison.

3.3.2.7) Summary of the EEG Statistical Analysis

A summary of the statistical results for the EEG analyses is shown in Table 3.7, which shows the P300 had the most significant main effects and interactions, whilst the N170 had the fewest. Table 3.7 also demonstrates there was less significant main effects and interactions within the auditory condition compared to the visual.

Table 3.7: *Summary of the results of the arithmetic task EEG statistical analysis. The ANOVA results for the mean amplitude, peak amplitude and peak latency for the N170, P300 and slow wave component across the analyses of the visual and auditory conditions. Main effects and interactions which reach significance p<0.001 are marked with **, main effects and interactions which reach significance p<0.05 are marked with *, and non-significant results are marked with 'ns'.*

		N170			P300			Slow Wave Component		
		Mean Amp.	**Peak Amp.**	**Peak Lat.**	**Mean Amp.**	**Peak Amp.**	**Peak Lat.**	**Mean Amp.**	**Peak Amp.**	**Peak Lat.**
Visual	**Difficulty**	ns	*	ns	ns	ns	ns	ns	ns	ns
	Electrode	**	*	ns	ns	ns	**	**	*	ns
	Hemisphere	ns	ns	ns	*	*	*	ns	ns	ns
	Difficulty x Electrode	ns	ns	ns	*	ns	ns	*	ns	*
	Difficulty x Hemisphere	ns	ns	ns	ns	ns	ns	*	ns	ns
	Electrode x Hemisphere	ns	ns	ns	ns	*	ns	ns	ns	*
	Difficulty x Electrode x Hemisphere	ns	ns	ns	ns	ns	ns	ns	ns	ns
Auditory	**Difficulty**	ns	ns	ns	ns	ns	ns	ns	ns	ns
	Electrode	**	*	ns	**	**	*	ns	ns	ns
	Hemisphere	ns	ns	ns	ns	*	ns	ns	ns	*
	Difficulty x Electrode	ns	ns	ns	ns	ns	ns	ns	ns	ns
	Difficulty x Hemisphere	ns	ns	ns	ns	ns	ns	ns	ns	ns
	Electrode x Hemisphere	ns	ns	ns	ns	*	ns	ns	ns	ns
	Difficulty x Electrode x Hemisphere	ns	ns	ns	ns	ns	ns	ns	ns	ns

3.4) Discussion

3.4.1) Mental Arithmetic Task in High-Specification Head-Mounted Display Virtual Reality

This study combined HS-HMD-VR with EEG recording equipment to successfully acquire ERP responses in a working memory VR arithmetic task. A 2x2 within-subject design was used, presenting ‘easy’ single-digit plus single-digit and ‘hard’ double-digit plus double-digit addition questions across visual and auditory modalities. ERP measures were recorded at the onset of the equation presentation, and behavioural measures were taken from the responses. WML-related N170 and P300 ERP components previously reported within the arithmetic literature were found across all conditions, and positive SWC ERP responses were present in the visual question presentations [131,239,241,244].

Statistical analysis of the ERP results find the visual ERP components largely correspond to the published literature. The visual N170 responses were largest in the central and parietal electrodes, with a larger peak amplitude for ‘hard’ questions [241,243]. Visual P300 responses were largest in the midline electrodes, and larger in the frontal region for ‘easy’ questions [204,223,238,239]. Within the visual SWC, it was found that the parietal region responses were more positive than the frontal response across both easy and hard conditions, coinciding with the typical positive SWC waveform [131,240,244].

Within the auditory trials, the auditory N170 was found to be larger in the central and parietal compared to the frontal electrodes. The auditory P300 responses were larger and earlier in the frontal and central electrodes compared to the parietal region. A larger auditory P300 peak amplitude was found in the midline electrodes compared to the right hemisphere, contesting the larger right-parietal response previously reported [223]. There were no significant differences between any comparison for the auditory SWC time-window.

The behavioural results further corresponded with previous findings, indicating that the harder questions evoked a larger WML than the easy questions based on measures of task accuracy, response time and subjective difficulty rating [219,229,230]. As far as the researcher is aware, at the time of writing this is the first study to utilise HS-HMD-VR in a full-head EEG ERP recording, or to examine mental arithmetic using HMD-VR.

3.4.1.1) Can Event-Related Potentials be Identified

This study aimed to investigate if ERP responses could be identified when utilising combined HS-HMD-VR and EEG recording equipment. The grand average ERP waveforms of the correct responses for both ‘easy’ and ‘hard’ question presentations in all 19 electrodes are shown for the visual condition in Figure 3.13, and for the auditory condition in Figure 3.14. The aim of this study was successfully achieved, with clear N170 and P300 components for both the visual and auditory presentations across both ‘easy’ and ‘hard’ questions being identified in each electrode recorded.

In the visual condition (Figure 3.13), a negative peak within the 120-200ms time window followed by a positive peak within the 200-400ms time window can be identified in all electrodes. The negative and positive peaks are identified as the N170 and P300 components respectively, corresponding with ERP components during arithmetic stimuli presentation reported in the literature [235,236,239,241,243]. The SWC can be identified by the post-400ms positive inflections in the majority of electrodes as the P300 peak observed slowly recedes to baseline over the course of the 1000ms. The SWC is further evidenced by several medial and right-hemisphere electrodes demonstrating a second positive inflection starting at the 500ms mark and peaking around 750ms. The visual SWC is most prominent in F8, but also present in F4, C4 and T4 [223,237,246], and to a lesser extend in Fz and Cz. However, post-400ms negative peaks can be seen in the left-frontal FP1 and F7 electrodes within the hard-visual condition.

Within the auditory responses (Figure 3.14), a negative peak identified as a N170 response within the 235-315ms time window followed by a positive peak identified as

a P300 within 200-400ms time windows can be identified. These peaks are identified as the N170 and P300 despite the later time windows used due to the auditory stimuli onset delay. The P300 is most prevalent in the central and frontal electrodes, particularly within Cz, C3, C4, Fz, F3 and F4. Examining the ERP waveforms for a post-P300 SWC as defined by parietal positivity and frontal negativity demonstrates no positivity in any parietal electrode. In all electrodes, the ERP response either returns to baseline or becomes negative to similar levels as the frontal electrodes, for example in Cz, C4 and P4.

3.4.1.2) Comparison with the Literature

Whilst ERPs were successfully acquired, HS-HMD-VR is a relatively new methodology within neuroscience research, particularly within mental arithmetic paradigms. It is therefore unknown if HS-HMD-VR usage introduces extraneous load or otherwise unexpected responses which render it unsuitable for WML research. To investigate if using HMD-VR results in unexpected outcomes in the WML arithmetic task, both the EEG and behavioural responses captured here were statistically analysed within presentation modalities. The results gathered are then contrasted against results reported in non-HMD-VR arithmetic studies to identify if the results gathered are congruent with the existing literature.

3.4.1.2.1) Behavioural

The visual and auditory behavioural results were consistent with the wider literature. When comparing between difficulties, it was found that responses in the hard trials across both modalities were significantly slower, less correct and received higher subjective difficulty ratings. Despite using different paradigms, previous research studying similar mental arithmetic tasks reported similar behavioural results in comparison between questions of different levels of difficulty. It is consistently reported across various mental arithmetic paradigms that increasing the difficulty of an arithmetic question results in longer response times and more incorrect responses [239,268,269]. The wider literature on the problem size effect within mental arithmetic also reports that longer response times and higher rates of incorrect responses are commonly found for tasks requiring a larger WML

[219,230,270]. Therefore, the behavioural results found in this present study suggests that load was successfully manipulated between difficulties within the current experiment.

Despite the similarities in behavioural findings, contrasting the current results against comparable studies using visually presented addition questions [268,269] (Figure 3.18) finds the mean response times are ~1s faster compared to the current results. In the current study, it was found that the mean response time was 2.6s for the easy-visual questions and 4.41s for the hard-visual, whereas Ashcraft & Kirk's [268] second experiment reported ~1.6s and 3.28s respectively for equivalent arithmetic trials requiring carries to be performed to reach the solution. However, the disparity does not indicate that the HMD-VR slows cognitive processing, and is instead argued to result from the differences between the paradigms. Whereas in the current study the participant must fully type out the response and press the enter button to submit the response, participants in prior studies only needed to press a single button [269] or speak the answer aloud [268]. This is reflected in the percentage of correct responses, where there are higher rates of correct responses in the current study, 99.13% and 94.13% for the easy-visual and hard-visual conditions in the current study, opposed to the 94.8% and 90.6% in Ashcraft & Kirk [268].

(a)

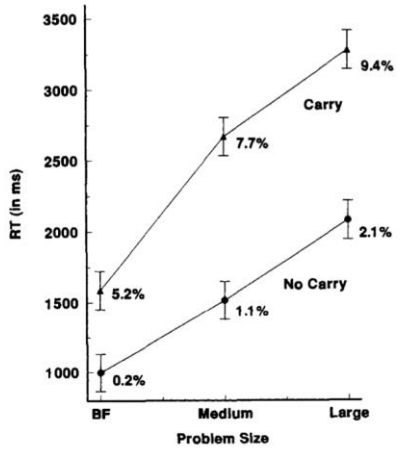


Figure 1. Mean reaction time (RT) and percentage error rates to basic fact (BF; *n* + *m*), medium (*nn* + *m*), and large (*nn* + *mm*) addition problems, separately for no carry and carry problems: Experiment 2. Error *bars display 95%* confidence interval based on *MSw*.

(b)

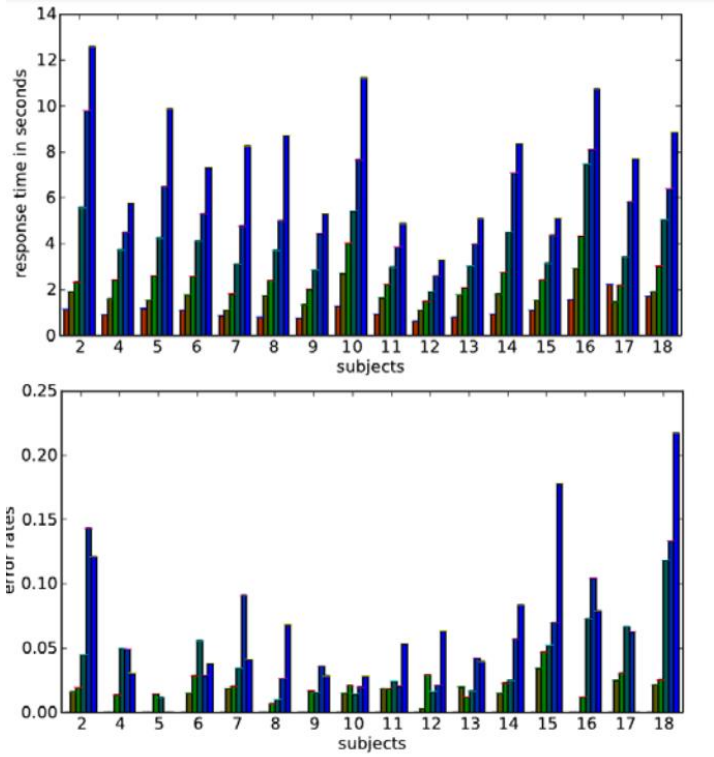


Figure 2: *Response time and error rate versus difficulty level for all subjects.*

Figure 3.15: *Graphs taken from (a) Ashcraft & Kirk [268] and (b) Rebsamen et al. [269] demonstrating the behavioural results of comparable conditions in mental addition tasks. For graph a, the 'BF' (one-digit plus one-digit) and 'Large' (two-digit plus two-digit) upper 'carry' bar is contrasted against the current study. For graph b, the first and middle bar presented for each participant represents the one-digit plus one-digit and two-digit plus two-digit trials respectively for response time (upper graph) and error percentage (lower graph).*

There were significant increases on the subscales of nausea, oculomotor disturbance, disorientation and total (cybersickness) score on the SSQ measured compared against an assumed rating of no symptoms. This is in line with other reported uses of different types of SB-HMD-VR and DB-HMD-VR devices [271]. However, only one participant of the twenty-two recruited exited the experiment due to discomfort, suggesting that degree of cybersickness-related symptoms introduced by the combined usage of HS-HMD-VR and EEG did not prevent the successful completion of the study.

3.4.1.2.2) P300

The most prominent component identified in the ERP waveforms evoked by the presentation of arithmetic questions is the P300, which is found in all electrodes included within the statistical analysis across difficulties in both presentation modalities. Mean amplitude analysis of the visually evoked P300 finds a significant interaction between hemisphere and difficulty, showing that the presentation of easy-visual questions evoked larger P300s in the frontal electrodes than for hard-visual questions; and that the central electrode has a significantly larger P300 than the frontal electrode in the easy conditions and the parietal electrode in the hard conditions. Within the mental arithmetic literature, the P300 component is commonly reported in response to the presentation of visual equations, auditory equations and potential solutions to equations across a range of task difficulties and problem sizes [131,223,237–239]. The finding of larger frontal P300s for the lower WML condition during the presentation of visual arithmetic questions is consistent with previous findings within the mental arithmetic literature [131,244]. Moreover, larger P300s have been reported in lower WML conditions within the wider working memory literature, for example in selective attention [272] and n-back [134] paradigms which compare tasks of multiple difficulties. However, the P300 found does not significantly differ outside of the frontal electrodes, nor was there enough differences between the easy-visual and hard-visual P300 for the main effect of difficulty to reach significance. The absence of a main effect of difficulty within the visual presentation of the question is also consistent with Dickson & Wicha [239], who reported no main effects of difficulty towards P300s evoked by the visual presentation of the second operand in a sequential presentation.

Moreover, the main effect of hemisphere was significant for mean amplitude, peak amplitude and mean latency analyses, which found the P300 in the midline electrodes were larger and earlier than those present in the right hemisphere, neither of which differed from the left hemisphere. This midline P300 is further demonstrated by the significant interaction between electrode and hemisphere within the peak amplitude analysis, demonstrating that the Cz electrode had larger P300 responses than the Fz and C4 electrodes, and that Pz P300 was larger than P4 P300. Within the literature, the P300 associated with mental arithmetic is primarily reported as

being largest in the left parietal region [131,236–238,273], opposed to the medio-central region found in the present study. However, many investigations reporting left P300 responses do not consider the midline electrodes [237,238,273], whilst studies which report the medio-central areas demonstrate the region is sensitive to different problem sizes in equation presentation [204,223].

Within the auditory condition, statistical analysis of the auditory P300s finds the main effect of hemisphere was significant within the mean amplitude, peak amplitude and peak latency analyses, finding that the parietal electrodes had a lower amplitude and later response than both the frontal and central regions, which did not differ. The peak analysis of the auditory P300 also revealed a main effect of hemisphere, showing that the midline P300 was larger than the right P300, neither of which differed from the left. Moreover, a significant interaction between hemisphere and electrode demonstrated the peak amplitudes of the parietal responses were lower than the central and frontal responses in the left, right and midline electrodes. These auditory findings directly contest Kiefer & Dehaene [238], who report that the auditory presentation of arithmetic stimuli resulted in the largest P300s being evoked in the right-parietal electrodes, opposed to there being the site of the smallest responses in the current study. Moreover, the lack of any significant main effect or interaction including difficulty contests the findings of Kiefer & Dehaene [238] and Dickson & Wicha [239], who reported that larger P300s are evoked by more difficult auditorily presented questions at the second operand.

The likely explanation for the auditory P300 findings is from the paradigm used in the current study. For at least the first 900ms, participants undergoing auditory trials would be attending to and subsequently encoding the equation, as opposed to having all the information and beginning the calculation within the visual condition. Moreover, as the first operand is still being presented during the early component time windows it is likely that little-to-no information about load was available to the participants at this time, explaining the lack of difference between difficulties. A potential alternative explanation is that the auditory ‘P300’ found is actually the P3a

as defined in Polich [274], who suggested that in auditory trials the P300 represents attention given to stimuli, and is found to be largest in the central electrodes.

3.4.1.2.3) N170

The earliest ERP component identified within the responses to presentation of the arithmetic questions is the N170. The statistical analysis of the visual N170 peak amplitudes finds significant main effects of difficulty, with the harder question presentation evoking larger responses. The increase in N170 amplitude towards the 'hard' visual trials correlates to the findings of Moore et al. [241] and He et al. [243], who reported similar increases in N170 amplitude to higher WML trials in arithmetic verification tasks. Whilst the increase in N170 amplitude for higher levels of WML is antithetical to findings within the face-perception literature [135], it is consistent with the limited arithmetic literature despite the differences between the production and verification paradigms employed.

Mean amplitude statistical analyses of the visual and auditory N170 revealed a significant main effect of electrode, which finds that the frontal N170 had a significantly smaller (i.e. less negative) amplitude than both the central and parietal region N170s across presentation modalities. Moreover, it was found that the frontal peak amplitude was significantly smaller than the parietal N170 peak amplitude in the visual condition, and than the central N170 in the auditory condition. However, no differences between hemispheres or any peak latencies were found. The auditory N170 did not differ between difficulties, likely due to the limited amount of information available to the participant about the task difficulty at the time the early ERP responses are recorded. No N170 analyses across either visual or auditory responses found a significant main effect or interaction including hemisphere, nor did peak latency differ on any comparison.

The N170 responses reported in this study largely coincides with arithmetic N170 reported in verification tasks [241,243], demonstrating that HS-HMD-VR can be utilised in experiments targeting early components. It is suggested that the N170

represents the initial encoding of the presented stimuli [241,243], which possibly explains the differences between difficulties in the visual condition and the lack of differences between difficulties in the auditory condition. Within the visual condition, there is a clear difference in how much information is presented to the participant at the point the N170 is recorded, with the hard condition having more digits presented at the same time. The N170 has previously been found to be sensitive towards the visual presentation of text-based words [275,276], thus it is reasonable to argue as there is more to encode the N170 would be larger. Conversely, the available information within the auditory condition at the point of the N170 recording is not enough to differentiate between difficulties, thus the 'amount' of encoding occurring between difficulties is balanced.

The arithmetic N170 response is often reported as having a left-hemisphere bias [241,243] that is not found in the current study. No potential explanation can be offered for this finding, as the N170 reported outside arithmetic stimuli is often largest in the right hemisphere [275,276], opposed to the lack of laterality found in the present study. Moreover, to the researcher's knowledge, auditory N170 responses to mental arithmetic stimuli have not been previously reported, nor are N170 responses often reported during speech perception tasks [277]. It is therefore unknown if the largest auditory N170 response being located in the central/parietal region is typical.

It is also possible that the 'auditory N170' reported is in fact an N1/N100 ERP response related to the identification and encoding of speech stimuli [278,279]. This is distinct from the N1 response reported for visual arithmetic stimuli associated with the identification of presented stimuli [222], which is centred around the ~200ms timepoint post equation onset. It has been reported that the latencies of visually-evoked and auditorily-evoked P300s using equivalent stimuli do not differ [280,281], thus it is reasonable to assume the P300 is time-locked between modalities. Therefore, the larger gap between the P300 and N170 time windows for the auditory condition compared to the visual implies the 'auditory N170' identified is an earlier negative component.

However, the difference in gaps between time windows between the auditory N170 and P300 is only 10ms (compared to the 0ms gap in the visual ERPs), and the previous comparisons of P300 stimuli were not conducted on arithmetic or numeric stimuli. This may imply the P300 is the P200 defined by Remijn et al. [278], however the prior findings of P300 in response to auditorily presented arithmetic stimuli [239] and actual onset of the auditory stimuli renders this unlikely. Therefore, a dedicated examination of the early negative components in response to auditorily and visually presented equations, specifically at the presentation of the first operand and second operand to examine any additional effect that WML may have on the components, would be beneficial to understanding neural response to arithmetic equations. Additional comparison between equations and syllable-matched three sentence words requiring true/false responses (e.g. ‘Rabbits are mammal’) could also allow identification of any effect of numeric stimuli compared to just speech.

3.4.1.2.4) Slow Wave

The latest component examined in this experiment is the post-P300 SWCs. Unlike the uni-directional N170 and P300, SWCs can comprise of increased parietal positivity, frontal negativity, or both, typically within the medial and right electrodes [131,223,244]. Visual inspection of the visual ERP waveforms gathered within the SWC time window shows that all responses are positive, tailing off from the P300 and persisting until 1000ms. Conversely, visual inspection of the post-P300 auditory ERP waveforms finds that the amplitude quickly returned to baseline or became negative in all electrodes.

Statistical analysis of the mean amplitude of the visual SWC finds a significant main effect of electrode, where both the parietal and central electrodes had a significantly higher amplitude than the frontal electrodes. Peak amplitude analysis also found a significant main effect of electrode, with the central peak being significantly larger than the frontal. There was an interaction between difficulty and electrode which found that the mean amplitude of the parietal and central electrodes had a significantly higher amplitude than the frontal electrode in the easy condition. Within

the hard condition, the parietal electrode had a significantly larger response than the central, which in turn was significantly larger than the frontal electrode. The significant interaction between difficulty and hemisphere found that the midline electrodes had a larger response than the left hemisphere during the hard question presentation only.

The increased positive response in parietal electrodes compared to the frontal electrodes aligns with the post-P300 positive SWC defined by Muluh [131] and the LPCs reported by Núñez-Peña & Honrubia-Serrano [246], Jasinski & Coch [237], Suárez-Pellicioni et al. [240] and Dickson & Federmeier [223]. Moreover, the hemispheric effect within the hard visual presentation suggests a midline bias similar to the results reported by Jasinski & Coch [237]. Whilst the hemispheric effect in the present study did not extend to the easy condition, nor between the often reported right hemisphere response, differences between hemispheres for post-P300 components are not always found [240,246].

It has been suggested that the SWC is associated with the mental calculation of the presented question, and as such has previously been reported to increase with larger problem sizes [131,248] and more complex operations [131,221] in response to visually presented arithmetic questions. However, the current visual SWC demonstrates no significant effect of difficulty. Whilst it has been previously reported the LPCs can be insensitive to load between different arithmetic operations [246], the similarity with Ku et al.'s [248] paradigm would suggest differences in difficulty would be expected during the different calculations. It is therefore possible that using HS-HMD-VR modulated experienced WML during the task which lead to the non-different result, however comparison between arithmetic tasks presented using HMD-VR and alternative display methods is required to examine this.

Statistical analyses of the auditory SWC finds no significant main effect or interaction within the mean amplitude and peak amplitude ANOVAs. A significant main effect of hemisphere is found within the peak latency analysis, but no post hoc comparison is

significant. As SWC have not been previously reported for the presentation of auditory arithmetic questions (to the researcher's knowledge), it is possible that the lack of difference should be expected. A potential explanation for this lack of difference is the auditory stimuli presentation itself. As stated, production tasks such as that utilised in the current experiment typically suggest the SWC represents the mental calculation process [131,248]. However, as calculation requires the presented question to be encoded, the lack of SWC within the auditory condition is likely due to the audio files of the question persisting into the 'SWC' time-window, thus the participant is still encoding at the point of the arithmetic process.

3.4.1.3) Event-Related Potential Conclusions

This study successfully acquired 3 ERP components associated with mental arithmetic processing during the visual and auditory arithmetic question presentation using HS-HMD-VR. There is no evidence to suggest that the use of HS-HMD-VR inhibited the EEG recording nor ERP analysis, demonstrating that HS-HMD-VR and EEG recording technologies can be used together in research contexts. Significant statistical differences between difficulty and electrode hemisphere and location were found within the ERP components, which correspond with responses reported in the wider arithmetic literature. Whilst some questions do remain about how the use of HS-HMD-VR influences the level of experienced load within WML paradigms, it is clear that HS-HMD-VR is suitable for EEG ERP and WML research paradigms.

3.4.2) Limitations and Future Directions

3.4.2.1) Comparisons with the Existing Literature

This study provides evidence supporting the use of HS-HMD-VR in neuroscience research. However, there are still differences in the ERP responses found that, when compared to the existing literature, highlight the limitations with the current experiment. As the behavioural results align with the wider literature, the differences in the ERP results are unlikely to result from the use of HS-HMD-VR. Instead, a more likely candidate is the difference between the current production paradigm and

verification paradigms used in many of the previous studies [223,237,240], or those targeting ERP responses to the presentation of the second operand [223,238].

3.4.2.2) Behavioural response time

It is possible that the response time behavioural measure does not accurately reflect calculation time differences between difficulties. Response time was measured from the response routine onset to when the 'Enter' button was pressed to end the routine. Therefore, it is possible that inputting responses to hard questions, which required 3-4 button pushes (2-3 digits and enter; M= 3.833s ± 0.0485 across all hard questions), resulted in increased response times relative to the 2-3 presses (1-2 digits and enter) required in the easy condition (M=2.567s ± 0.0645 across all easy questions). However, the average time for the easy condition for both modalities was ~2.5 seconds, which if the participant knew the answer prior to input onset would mean each button press requires ~1s to input. If only an additional 1.3s button press is required for the hard condition, this would not account for the 2-5 second difference between difficulties within conditions. A more likely explanation for the differences between conditions is that the hard questions had longer responses due to inducing a higher level of WML [239]. A potential method of verifying this was explored based on the first number button press on the input screen (opposed to the enter button) for correct responses. However attempts to extract the time of the first button press was unsuccessful due to an issue with the experiment software not consistently logging all button presses in the output file, opposed to the routine time of the input response screen.

3.4.2.3) Limited Differences Between Task Difficulties

A potential limitation of the current study is the relatively low amount of significant differences found between difficulties when examining the neurological data. Whilst this study achieves its aims and acquires ERP responses, the main effect of difficulty was only significant for the N170 peak amplitude, and many interactions did not reach significance. The lack of P300 differences is particularly surprising, as the component is noted for being sensitive to problem size in working memory arithmetic tasks [223,239]. A potential explanation for the limited differences is the number of

participants included in the analysis, as increasing the number of participants, and in turn the number of trials included in the ERP analysis, can increase the chance of finding statistically significant findings [282]. However, as previously stated, using ~20 participants is standard for EEG research (see Larson and Carbine [183]), with particularly large responses such as the P300 requiring as few as 5-7 participants [283]. Moreover, significant EEG comparisons within similar arithmetic tasks to that used in the current study were found by studies using between 8-10 participants [227,229,235]. The minimum threshold of epochs per condition, suggested to be at least 6 epochs, required for analysis was also exceeded for all participants included in the final analysis [282,284]. Whilst it is possible that recruiting more participants would increase the number of significant differences found between difficulties within the EEG comparisons, it is unlikely that it would change the overall conclusion of this research, nor does it diminish the findings of the present study.

A more plausible explanation for the limited differences between neurological measures of WML is that the difficulties between the ‘easy’ and ‘hard’ conditions was not as large as the behavioural results may suggest. It is possible a ceiling effect resulting from the ‘easy’ single digit equations and the relatively low-difficulty ‘hard’ condition exists, which may limit the differences in neurophysiological responses measured in the current study. The average number of errors is <1 in both easy conditions, <4 in the hard visual condition, and <5 in the hard auditory condition. It could therefore be argued that participants performed well across difficulties and presentation modalities, and that the relative ease of which participants solved ‘hard’ equations did not result in large increase of experienced WML. Moreover, studies with similar comparisons between digit sizes or q-values have referred to the difficulty categories used in the current study differently. For example, in both Spüler et al. [227] and Chin et al. [229], the ‘hard’ questions in the current study would fall into the “medium” category. Moreover, Ullsperger et al. [235] refers to single-digit equations as “very easy” and the double-digit equations as “easy”, and reported larger differences in EEG responses were found when using “medium” three-digit equations.

By increasing the difficulty of the ‘hard’ conditions, for example by increasing the q-value bracket of the equations [229] or using an equation with a different operation [244], a larger distinction between conditions can be identified which could increase the power of the results found. Such comparisons would in turn provide additional insight into how HMD-VR interacts with experienced WML, and would be particularly beneficial when compared against other display methods to identify the advantages and limitations of HMD-VR presentation. Increasing levels of WML by raising the values of the highest Q-value bracket used may also produce more significant comparisons involving difficulty within the statistical analysis, increasing the power of the current results and providing additional insight into neurophysiological responses to mental arithmetic tasks. However, increasing the WML experienced in the context of the present study is not expected to change the conclusions drawn, as it is unlikely that increasing task difficulty would prevent the acquisition of ERP responses.

3.4.2.4) Better implementation of the SSQ

The present study employed the SSQ at the conclusion of the experiment to identify if participants experienced adverse effects from combined HS-HMD-VR and EEG usage. However, as the SSQ was not also taken before the HMD-VR was placed on the participant, it is possible that participants were experiencing some minor level of cybersickness-related symptoms prior to onset due to reasons unrelated to the HMD-VR device. It is therefore recommended that future HMD-VR experiments utilise both a pre- and post-HMD-VR exposure (and any other display) comparison of cybersickness to identify any effect that HMD-VR has on cybersickness symptoms, as has been employed by Sharples et al. [285] and Xu et al. [181]. It is also unknown how the combination of EEG and HMD-VR may further increase discomfort, for example due to additional weight on the head or due to the HMD-VR head straps applying pressure to the EEG electrodes against the scalp. Therefore, a comparison of the SSQ scores between different combinations of EEG and HMD-VR systems is also suggested to identify how using the technologies can additionally increase participant discomfort, as to minimise any negative affect in future research.

3.4.2.5) Comparison with Other Display Methods

The successful acquisition of ERP components coinciding with those reported in the literature is an important step for the utilisation of HS-HMD-VR in neuroscience research. However, this study does not examine how HS-HMD-VR influences experienced WML compared to lower-immersion display methods. Moreover, direct comparison between the amplitudes of ERP components acquired in the present study and those reported in the literature cannot be conducted due to the differences in paradigms utilised. The task used in the present study utilises a production task which measures ERP responses from the presentation of the whole question, which differs from the onset of the second operand in other production paradigms [238,239] or solution onset in verification tasks [223,237]. Understanding how the use of HS-HMD-VR modulates WML is the next step to understand how to best utilise the technology in psychological and neuroscience research. Therefore, comparison between HS-HMD-VR and an alternative display method such as DB-VR utilising the same task should be conducted, allowing for the direct comparison of differences in experienced WML between conditions.

3.5) Conclusions

The results found in this current study have important implications for the use of HS-HMD-VR in wider neuropsychological research. The results found demonstrate that EEG signals can successfully be acquired when using HS-HMD-VR equipment. The behavioural results have demonstrated that WML can be successfully manipulated within HS-HMD-VR devices. Moreover, the EEG results for the visual conditions found largely that lower WML arithmetic questions induce larger P300 components and similar response patterns within the SWCs. These results indicate that the advantages offered by HS-HMD-VR in research can be utilised without inherently introducing overwhelming WML nor levels of noise that prevents the successful acquisition of neurological responses.

COVID-19 Impact Statement

In early 2020, the COVID-19 global pandemic prevented access to the EEG laboratory for an extended period of time. Following the reopening of the laboratory, the requirements of masks and face shields made the collection of combined HMD-VR and EEG data impossible for several more months until restrictions were lifted. Tests of a standard medical mask and a double-layered fabric mask with HMD-VR resulted in the HMD lenses repeatedly fogging up. The fogging up persisted over several removals, cleanings of, and reapplication of the HMD-VR device. Whilst this may be resolvable in certain circumstances, the fogging ran a very high risk of interrupting any experiment, being uncomfortable to participants (particularly those with glasses), and shifting the EEG electrodes when cleaning the lenses. Moreover, an uncomfortable level of heat from combined face mask, EEG headcap, and HS-HMD-VR device was anecdotally reported from members of the laboratory who experienced the combination. To prevent participants experiencing unnecessary cybersickness symptoms, or otherwise increasing sweat-related artifacts to the recorded EEG signal, data collection was paused until the mask restrictions were lifted.

I made efficient use of the duration where the COVID-19 restrictions prevented data collection by improving sections of this project. Firstly, the systematic review was updated to include the most recent papers, and then extended to include additional working memory papers as discussed in Chapter 2. Moreover, several time-frequency analysis methodologies were prepared for the (at the time) planned second EEG experiment were designed and coded in Matlab. These methods included event-related spectral perturbation (ERSP), event-related desynchronisation/synchronisation (ERD/ERS), and power-spectral analysis methods, and were selected to facilitate a range of potential working memory paradigms for use in HMD-VR.

The following chapter will discuss the re-analysis of the Chapter 3 arithmetic data in the context of different preprocessing decisions used in HMD-VR research. During

the analysis of the arithmetic data, it became apparent that there was little consistency between preprocessing methodologies used in published HMD-VR/EEG data. There is minimal published research utilising the HS-HMD-VR Vive Pro device used in the arithmetic study in combination with EEG, making it difficult to identify how the EEG artifacts identified in the arithmetic data were best removed. Therefore, a dedicated exploration of different EEG preprocessing pipelines was compared using actual HS-HMD-VR/EEG data collected in the arithmetic study, with the aim of ensuring appropriate preprocessing decisions were made in this thesis.

Chapter 4) A Comparison of Preprocessing Steps for Artifact Removal in Event-Related Potential Electroencephalography Data Collected During a High-Specification Head-Mounted Display Virtual Reality Arithmetic Task

4.1) Introduction

The underlying principles of EEG as a neuroscience technique have remained consistent since the inception of the method in 1924 [48]: the brain produces excitatory or inhibitory activity in response to stimuli or tasks that can be measured as electrical signals. The signals generated by the brain can then be compared against another condition or baseline to isolate the neural response to the stimuli. Between the EEG recording electrodes and the brain is layers of skull, scalp, meninges, and cerebrospinal fluid, necessitating sensitive electrodes and amplifiers to enhance the signal to a measurable level [286]. However, the signal enhancement also renders EEG sensitive to non-brain physiological and environmental sources of electrical noise, which can contaminate the recorded data and subsequent ERP analysis if improperly processed [286]. Despite concerns regarding the proximity of an HMD-VR device to the recording electrodes, Chapter 3 has demonstrated that HS-HMD-VR and EEG can be successfully combined to acquire ERP responses. During the analysis process conducted in Chapter 3, three types of artifacts were found and successfully minimised or removed through preprocessing steps: eye movements and eyeblinks; line noise; and slow-wave drift.

Line noise at the 50/60Hz frequency and associated harmonics can be introduced to EEG recordings due to the proximity of the EEG cap to electronic devices [56,287]. Slow-wave drift, commonly associated with increased perspiration [288,289] affects electrode impedance by introducing a gradual change in the amplitude of a recorded signal over time [290,291]. Sweat-related slow-wave drift may have resulted from the increased temperature within the HMD-VR increasing perspiration [176,292] or

increasing symptoms of cybersickness [194,293]. The increased temperature may also cause the conductive electrode gel to dry quicker, changing the electrode impedance [294]. Ocular artifacts caused by eye movements, such as eyeblinks, were prominent in many datasets captured in the Chapter 3 arithmetic experiment. Eyeblinks have previously been associated with symptoms of cybersickness [295,296], however they may also result from individual differences between participants [297].

The artifacts identified can be reduced or removed using appropriate preprocessing steps to reduce the amount of noise. Highpass filters are used to remove slow drift [298], line noise and harmonics can be removed by lowpass or notch filtering [56,287], and eyeblinks can be removed using epoch rejection [205] or ICA [299,300]. However, incorrect selection of preprocessing parameters can distort the data and reduce statistical power of extracted ERPs [298]. Moreover, the preprocessing parameters selected also depends on the data analysis method planned. Certain preprocessing steps are better suited for specific analytical purposes, such as improving data quality [301] or facilitating specialised statistical analyses [302], meaning comparable paradigms are often preprocessed differently. For example, Brouwer et al. [303] filtered between 0.1-100Hz without artifact rejection for online classification during a n-back task, whilst Scharinger et al. [134] filtered between 0.5-40Hz and used ICA-based ocular artifact rejection to facilitate ERP and time/frequency analysis for a comparable n-back paradigm. Despite the differences, both studies successfully reported P300 responses, demonstrating that there is not a singular ‘correct’ way to process data. Instead, preprocessing must balance optimising the data for analysis with removing noise from the EEG recording. Therefore, it is important to understand how each preprocessing decision changes the EEG data.

4.1.1) Introduction to Main Electroencephalography Processing Steps

4.1.1.1) Filtering

Filtering is a common early step in EEG preprocessing pipelines, and is used to remove artifacts such as line noise and slow drift by modulating targeted frequency

bands [304]. Filters can be divided between categories, which define the frequencies targeted relative to a given value or values. Categories of filters include: highpass filters, which reduce the amplitudes of frequencies below a defined value; lowpass filters, which reduce the amplitudes of frequencies above a defined value; notch filters, which reduce amplitudes of frequency between two values; and bandpass filters, which reduce the amplitude of frequencies outside of two values. Filters are used to target certain artifacts in EEG recordings, for example using either a 40Hz lowpass, a 49-51Hz notch filter, or a 0.1-40Hz bandpass filter would remove 50Hz electrical line noise. A 51Hz highpass filter would also remove the 50Hz noise, however it would also remove most of the data which contributes to the ERP trace.

Filters are also divided between types based around ‘response functions’, which define how the targeted frequencies are modulated or attenuated. Filters used in EEG gradually attenuate the frequency around a ‘half-amplitude cut-off’ [298] where the signal is attenuated to half its original strength. The order of the filter determines the range of surrounding frequencies that are attenuated, and how much attenuation is applied. Higher order filters affect a wider range of surrounding frequencies, but gradually increase attenuation as the half-amplitude cut-off is approached. Lower order filters are ‘steeper’, affecting fewer surrounding frequencies but having larger increases in attenuation between each included frequency.

Frequency response function (FRF) filters apply the response function in the frequency domain. First, the EEG data is Fourier transformed into the frequency domain where a sloped FRF filter with a gain between 0-1 is applied centred on the half-amplitude cutoff. The slope’s direction is determined by the category of filter used and the steepness is defined by the filter order. Filters targeting two frequencies will be comprised of two slopes. The power of each frequency is multiplied by the corresponding gain, and then transformed back into the time domain using the inverse Fourier transform. Conversely, impulse response function (IRF) filters are applied to time domain data through a process called ‘convolution’. Each data point within the time range (the impulse) is replaced by the IRF scaled to

the amplitude of the impulse it is replacing, and averaged together with a number of surrounding datapoints dependant on the order of the filter.

IRF filters can be further divided between finite impulse response filters (FIRs) and infinite impulse response filters (IIRs). FIRs are ‘causal’, only including previous time points when applying the filter. However, causal filters can result in phase distortion causing a temporal shift in the data [305–307], moving the peaks in ERP analysis and potentially leading to the misanalysis of the signals [308]. There is also the risk of ‘edge artifacts’, where the filter applied at the start and end of the dataset is improperly applied due to the lack of available timepoints [298]. FIR filters have been used in combined HMD-VR and EEG experiments [66].

IIR filters such as the Butterworth filter are ‘acausal’, using both forward and backward filters to account for the ‘future data’. IIR filters are ‘zero-phase’ and do not introduce the phase delays [298] whilst maintaining the shape of the waveforms more accurately compared to causal filters [309], reducing the risk of misanalysis due to filter artifacts. Butterworth IIR filters are often reported in combined EEG/HMD-VR studies [179,205,287], and were utilised in the Chapter 3 arithmetic experiment.

The filter category and parameters also have a large effect on the ERP waveform [310]. Inappropriate selection of filter values can result in noise not being suitably removed, or desirable frequencies being over-attenuated resulting in reduced amplitude [298]. To date, no research comparing the effects of various filtering parameters on data collected from a combined EEG and HMD-VR study has been conducted. Therefore, it is important to identify what filtering parameters have been used in the existing literature.

4.1.1.2.1) Highpass Filtering

Highpass filters attenuate frequencies under the defined value, and are therefore used to remove slow wave frequencies, for example sweat-related drift, from the EEG data [291]. Hypothetically, the upper limit of a highpass filter is the minimum frequency a researcher is interested in, which in EEG analysis is typically 1Hz [311]. However, Tanner et al. [310] demonstrated higher filter cut-offs can reduce the amplitude of ERP components (Figure 4.1), and reduce the power of statistical analysis. Highpass filtering therefore requires balancing the artifact removal without over-attenuation [307,310].

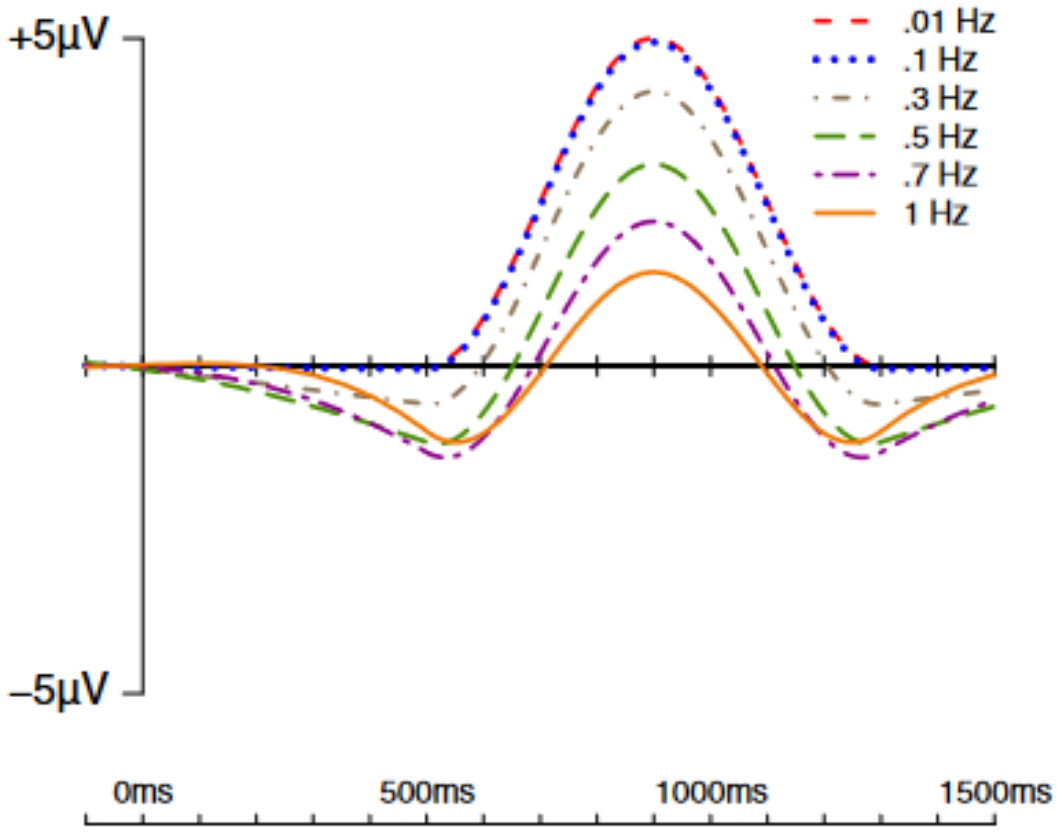


Figure 4.1: *The effects of 0.1Hz to 1Hz highpass filter values on a simulated P600 component, taken from* Tanner et al. [310]*. This figure displays how selecting a higher half-amplitude cut-off frequency during high-pass filtering affects the amplitude of extracted ERP components.*

An upper limit of 0.1Hz for highpass filters is often recommended in non-HMD-VR research, as it offers the ‘optimal trade-off’ between noise reduction and potential filtering artifacts without over-attenuating the peaks in the ERP waveform [309,310]. In the context of HMD-VR usage, the main consideration is countering the sweat-related drift. Aksoy et al. [205] report that HMD-VR/EEG studies have used anywhere between 0.1-3Hz, though typically 0.5Hz and above [16,25,205]. As the reasoning for the selected frequencies is not often stated, it is uncertain if the higher filter value is to remove sweat-related drift, or to serve another purpose. Therefore, a comparison of high-pass thresholds for removing slow-drift artifacts will be conducted to identify which are suitable for combined HS-HMD-VR and EEG research.

4.1.1.2.2) Lowpass and Notch Filtering

Lowpass filters attenuate frequencies above a defined threshold, which can be used to remove high frequency noise caused by electronic interference [311]. However, lowpass filters can distort ERP responses by reducing the amplitude of averaged waveforms, or introduce uniformed temporal shifts to the onset and offset of waveforms [298]. An alternative option is notch filtering, which targets and reduces specific frequencies bands, for example between 49-51Hz to remove 50Hz line noise. Notch filtering is commonly applied in EEG preprocessing [312], as unlike lowpass filtering it does not remove higher frequency bands that may be of interest to researchers. Notch filters typically have a high order, and thus remove artifacts without attenuating the surrounding data. However, like lowpass filters, notch filters can introduce phase shift or other time domain distortions when improperly applied [312]. As a result of the concerns of using lowpass and notch filters, some researchers suggest not using a lowpass filter at all [313], or only using them when presenting the ERP data for additional clarity [307]. However, using no high-frequency filtering can result in electrical noise introduced by an HMD-VR device remaining in the data.

Whilst it may seem obvious to remove the high-frequency noise from the data due to the proximity of an electrical device to the EEG electrodes, there are examples of no

lowpass or notch filtering being applied in HMD-VR studies [205]. Therefore, it may suggest that any line noise introduced by HMD-VR is negligible for ERP analysis. However, other HMD-VR papers have reported using lowpass filters ranging between 40Hz [16] to as low as 20Hz (via a bandpass filter) [25]. Notch filters have also been applied within WML-based HMD-VR studies [64] and motor-tasks [314]. Due to the range of potential high-frequency filters available, comparison between methods used to remove (or not remove) high-frequency noise would be beneficial to understanding how HS-HMD-VR and EEG can be combined in research applications.

4.1.1.2) Eye Movement Artifact Rejection

EEG is often contaminated with physiological noise, with eye movements, mouth movements, swallowing, muscular activity, and heartbeat contamination introducing sudden spikes to the EEG waveform [315]. The most common artifact identified in the Chapter 3 experiment was eyeblinks, which comprise of a distinct increased/decreased amplitude pattern that can be seen in the frontal electrodes extended posterior along the scalp (Figure 4.2) [316]. As the eyelids open and close and contact is made with the cornea, or lateral and horizonal eye movements are made, the cornea moves. As the cornea is a dipole, changes in its orientation or conductance produce detected signals in the EEG recording [315,317]. In ERP studies, sudden increase in amplitude can result in false peaks in the averaged waveform, which in turn can lead to misinterpretation of the results [317]. Moreover, if HMD-VR does increase eyeblinks in certain participants [295,297], the chance of overlap between averaged epochs is exacerbated.

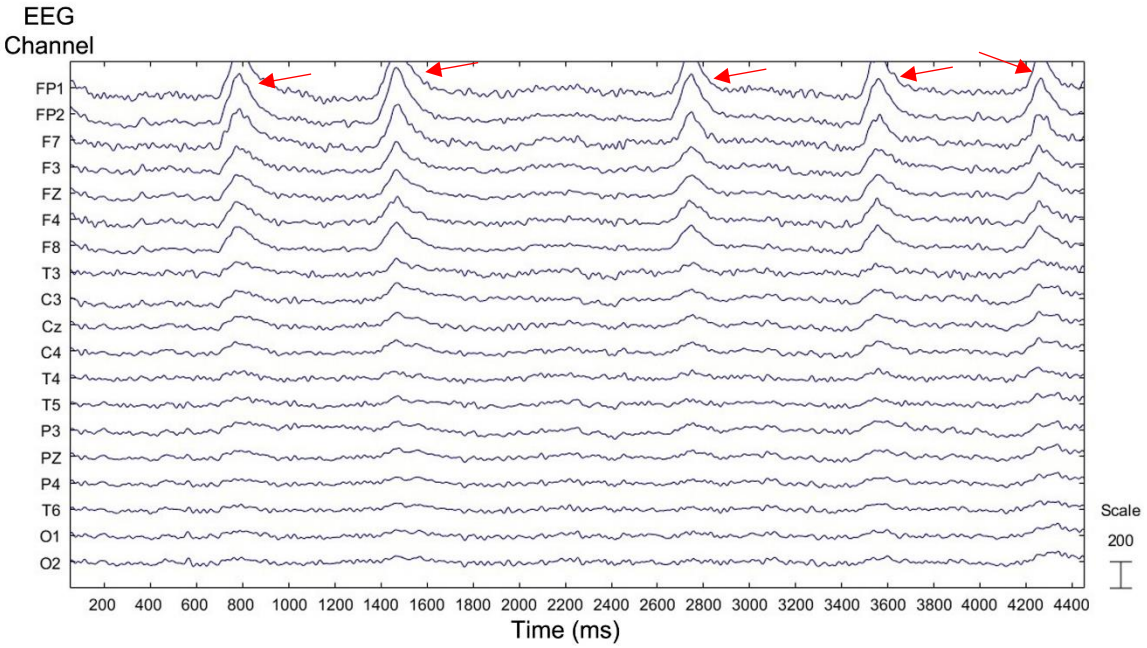


Figure 4.2: *Example of 5 eyeblinks in EEG data, taken from the arithmetic study data (see Chapter 3). Eyeblinks are marked with arrows (red colour). The data is filtered between 0.5Hz and 40Hz for clarity. The eyeblink artifact is most prominent in the frontal electrodes, and reduces in amplitude towards the posterior electrodes.*

One way of removing eye movement artifacts is through the rejection of affected data epochs, either manually through visual inspection or using an automated detection algorithm, such as the threshold detection in EEGLab [261]. Whilst rejecting contaminated epochs has been utilised to remove artifacts in HMD-VR studies, removing too much data can potentially leave insufficient data for analysis [282,284]. An alternative method is Independent Component Analysis (ICA), which is an algorithm that estimates the distinct sources of signals which comprise the complete EEG waveform [318–320]. As EEG data recorded at each electrode constitutes the combined activity of distinct but connected areas of the brain (and other sources of noise), it is possible to estimate each distinct contribution ('source') to the recorded signal to form a 'mixing matrix' that shows how they interact across the scalp. Based on the assumption that the sources are truly independent and non-gaussian (not normally distributed), ICA identifies and separates ('decomposes') an estimate of these sources into an 'unmixing matrix' by finding 'maximally distinct/independent' ICA components, meaning there is no overlapping information between identified ICA components. The unmixing matrix generated separates the

scalp data by channels-by-components, therefore the number of outputted components will typically be the same as the number of EEG channels included. ICA components can include responses to the stimuli, unrelated processes occurring within the brain, and bodily or external artifacts which introduce noise into the recorded data. The contaminated ICA components can be selectively removed before rebuilding/recombining the data, excluding the artifact whilst preserving the brain activity recorded.

It has been found that ICA is highly proficient at removing eyeblinks without altering surrounding data [300], whilst comparable methods such as principal component analysis can alter the amplitude of surrounding non-blink data [299]. ICA has been used in HMD-VR experiments to remove eyeblink artifacts [25,167]. However, ICA is a non-linear operation which produces slightly different results each time decomposition is performed (Figure 4.3). Pontifex et al. [284] compared P3 amplitudes of the same data, processed several times within and between different ICA algorithms to remove eyeblinks from the data. They found the amplitude of the P3 ERP component varied within the same ICA parameters, and between the different ICA algorithms. Therefore, care must be taken when utilising ICA methods to not reduce the power of the recorded components.

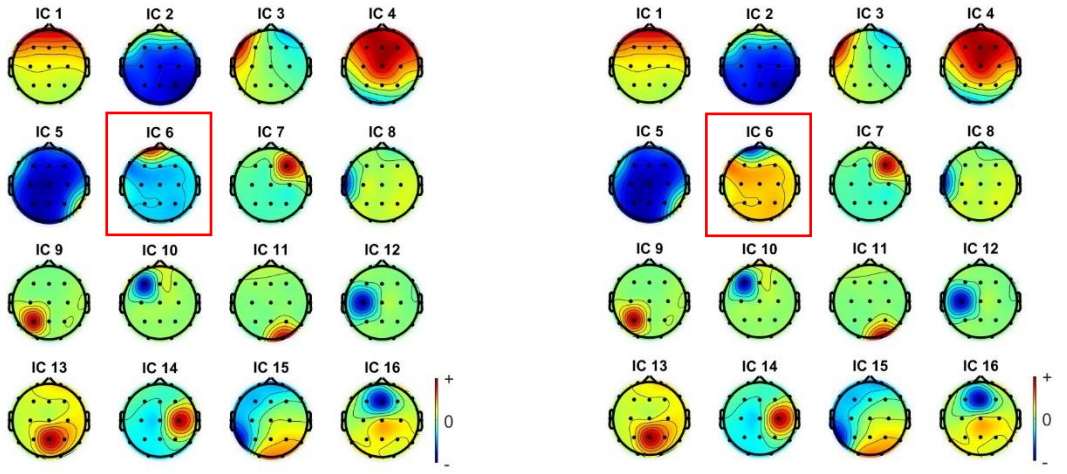


Figure 4.3: *ICA components identified in the same dataset by two runs of a RUNICA algorithm in Matlab. The data is from a single participant in the arithmetic study (see Chapter 3). Component 6 (highlighted) is reversed in polarity between the examples, despite the parameters being identical.*

The question remains of which method is optimal for removing eye-based artifacts from ERP datasets collected during HMD-VR use. As both improperly implemented ICA and epoch rejection risks reducing the power of ERP components, a comparison between the methods to identify which is suitable for HS-HMD-VR/EEG experiments is required. Comparisons between the methods should include how they affect the recorded ERP waveforms, the amount of data removed from the dataset, and the changes within the ERP peaks identified in Chapter 3.

4.1.2) Aims of the Present Study

There is limited guidance or consensus on which preprocessing parameters should be utilised when analysing EEG data collected when using HMD-VR. As the use of HMD-VR, particularly HS-HMD-VR, is still relatively unexplored in the EEG literature, it is important to identify how to avoid errors in the preprocessing selections that may negatively affect the data.

The aim of this chapter is to examine how common preprocessing steps in EEG analysis can be used to remove artifacts found in EEG recordings during HS-HMD-VR. Unprocessed data collected for the Chapter 3 arithmetic experiment, which contains eyeblinks and eye-movement artifacts, 50Hz electrical line noise, and slow drift artifacts, will undergo different variations of steps in preprocessing pipelines. The three preprocessing steps varied are 0.1Hz highpass filter (0.1HzHp), 0.5Hz highpass filter (0.5HzHp) and 1Hz highpass filter (1HzHp) for slow drift removal; No lowpass (NoLp/NF), 50Hz notch filter (50HzNF), and a 30Hz lowpass filter (30HzLp) for high frequency electrical noise, and no eye-based artifact removal (No-EBAR), ICA eye-artifact removal (ICA-EBAR), and epoch-rejection eye-artifact removal (Epoch-EBAR) for eyeblink artifacts. Visual inspection of the continuous EEG waveform and comparison within the N170 and P300 peak and mean amplitudes of the Cz electrode are considered when comparing between preprocessing parameters. The P300 and N170 components within the Cz electrode were selected for examination as these were the largest responses identified in Chapter 3 arithmetic study.

4.2) Methods

4.2.1) Equipment and Procedure

The participants, materials, apparatus, and experimental procedures used are detailed in the Methods section of the arithmetic study in Chapter 3.

4.2.2) Electroencephalography Data Processing

The data analysis conducted on the ERP data utilises a modified version of the 3-stage preprocessing pipeline defined in Chapter 3, which varies at three key stages to facilitate comparisons within highpass filter values, lowpass/notch filter parameters, and eye movement component removal methods. The EEG processing was conducted using the EEGLAB toolbox [261] for MATLAB [262] including the ERPLAB plugin [263].

All datasets undergo the same first four steps, where the data is imported into a file format compatible with EEGLab, non-EEG channels removed, and the electrode location data inserted. The first branch of the analysis is the highpass filter, where datasets either had a 0.1Hz, 0.5Hz, or 1Hz filter applied. The highpass filter is followed by either a 30Hz lowpass, 50Hz notch or no additional filter. Channels requiring interpolation are identified by applying a temporary 30Hz lowpass filter, which are then interpolated on the non-temporary filtered data.

Each analysis variant undergoes separate ICA preparation and processing in stage 2. To extract eye movement components, a 30Hz lowpass filter is applied to the no lowpass and notch filtered datasets, epochs are extracted between -200 to 4504ms, and epochs containing non-eyeblink noise are rejected. The ICA algorithm is then applied to each dataset, and the ICA weights are saved.

Returning to the datasets generated at step 7 following channel interpolation (Figure 4.4), -200 to 4504ms epochs are extracted and the respective ICA weights for each variant are applied. Due to the number of variants, the ICA components for each

dataset are automatically classified using the ‘label ICA components’ function, which examines the hemispheric origin and power of the extracted components to identify if the component results from the brain, eye, heart, muscle or other. Once the components are identified, datasets either have no eyeblinks removed, eye-related components removed using ICA, or eyeblink artifacts removed using epoch rejection, resulting in 27 total variants for each dataset. Noisy epochs were identified using the pipeline with the highest frequency value for the highpass filter (1Hz), the lowest value for the lowpass filter (30Hz Lowpass), and the ICA-EBAR method for eye movement removal to minimise the chance data removed resulted from any of the three targeted artifacts. The epochs were then rejected from all pipeline variants for consistency in the data compared. In the Epoch-EBAR condition, the same epoch rejection process was conducted to remove epochs contaminated with eyeblinks within the -200 to 1000ms time window, and applied to all Epoch-EBAR variants. The average ERP waveform for each individual dataset is then calculated, then combined within each variation of the pipeline together to produce the 27 grand average datasets.

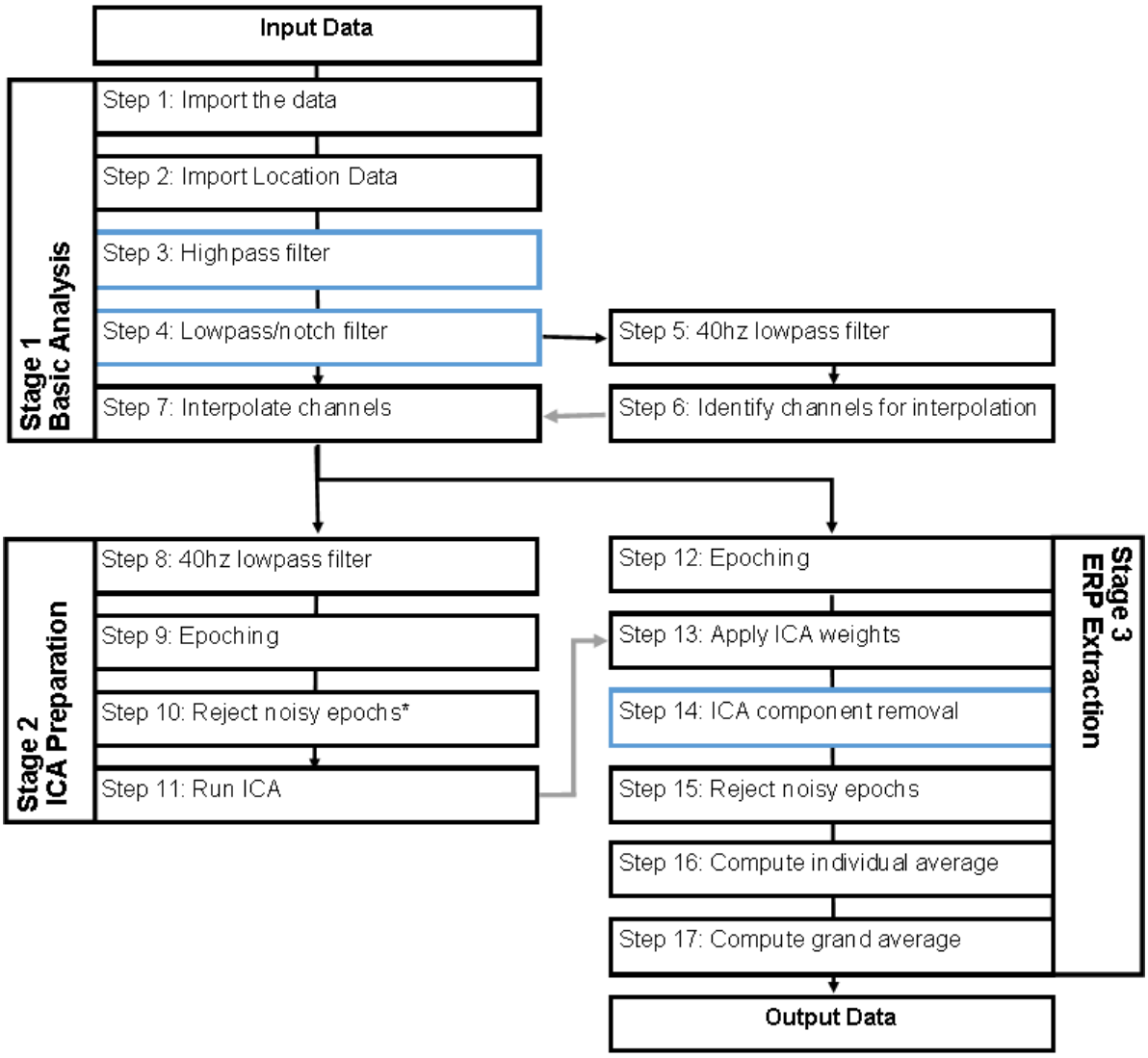


Figure 4.4: Diagram of whole processing pipeline numbered by the order of processing steps performed on the data. Black arrows show where the entire dataset is used in the connected step. Grey arrows show where only part of the information is used from the connected step. Parts of the pipeline which vary for comparison are highlighted with a blue outline. Step 10 (marked with an *) only rejects epochs containing non-eyeblink noise.

Level of line noise is also extracted by isolating the 200ms baseline for each epoch. The baseline period is then converted into power spectral density using a fast Fourier transform to identify the level of 50Hz line noise contamination.

4.3.2) Statistical Analysis

Statistical analysis was conducted in SPSS 28 [264] and Microsoft Excel [265]. Four separate three-factor 3x3x3 [highpass (1HzHp, 0.5HzHp, 0.1HzHp) x lowpass

(30HzLp, 50HzNF, NoLp/NF) x ICA (No-EBAR, ICA-EBAR, Epoch-EBAR)] repeated measures ANOVA were conducted to compare the effect of the different processing variations on the outputted data. The first and second ANOVAs targeted the peak and mean amplitudes of the N170 component within the Cz electrode, and the third and fourth ANOVA targeted the peak and mean amplitudes of the P300 component within the Cz electrode. Mauchly's test of sphericity was used to identify any main effect or interaction which violated the sphericity assumption. When sphericity violations are identified, the degrees of freedom are adjusted using the Greenhouse-Geisser or the Huynh-Feldt corrections when the Greenhouse-Geisser Epsilon is under 0.75 or over 0.75 respectively. Post-hoc analysis was conducted using pairwise comparisons, corrected for multiple comparisons using the Bonferroni correction to control for type 1 error [267].

A paired-sample t-test was used to compare the number of removed epochs between the No-EBAR/ICA-EBAR conditions and the Epoch-EBAR conditions to identify if significantly more epochs were removed in the Epoch-EBAR variants.

4.3) Results

4.3.1) Epoch Rejection

A paired-sample t-test found that there were significantly less epochs included in the Epoch-EBAR conditions (M=90.714± 9.973) from the initial 240 per participant compared to the No-EBAR/ICA-EBAR conditions (M=201.762± 2.772) ($t(20)=10.347$, $p\leq0.001$, $d=2.258$). The same number of epochs were removed in the No-EBAR and ICA-EBAR conditions.

4.3.2) Effect of Highpass, Lowpass and Eye-Based Artifact Rejection on Continuous Data and Topographic Event-Related Potentials

4.3.2.1) Peak and Mean Amplitudes

Within the N170 responses, using 0.1HzHp, NoLp/NF and Epoch-EBAR results in the largest peak (M=-15.9 ± 3) and mean (M=-6.7 ± 2.1) amplitudes. Using 0.1HzHp,

a 30HzLp and ICA-EBAR results in the smallest peak amplitude (M=-3.1 ± 0.9). The smallest N170 was found using 0.1HzHp, a 50HzNf and ICA-EBAR results in the smallest mean amplitude (M=-0.6 ± 1).

Within the P300 responses, using 0.1HzHp, NoLp/NF and ICA-EBAR results in the largest peak amplitude (M=21.2 ± 1.6). The largest mean amplitude resulted from using 0.1HzHp, 50Hz notch filter and ICA-EBAR (M=13.1 ± 1.2). The smallest peak amplitude is found when using 1HzHp, 30HzLp and No-EBAR (M=10.5 ± 1), and using 1HzHp, no lowpass and No-EBAR results in the smallest mean amplitude (M=4.4 ± 0.8).

4.3.2.2) Highpass Filtering

Visual examination of the ERP waveforms comparing highpass filtering level from all recorded electrodes (Figure 4.5) shows that using a 0.1HzHp filter results in a positive inflection in the Fz, F4, F8, C3, Cz, C4, T4, T5 and Pz electrodes starting at ~300ms and persisting until the end of the 1000ms waveform.

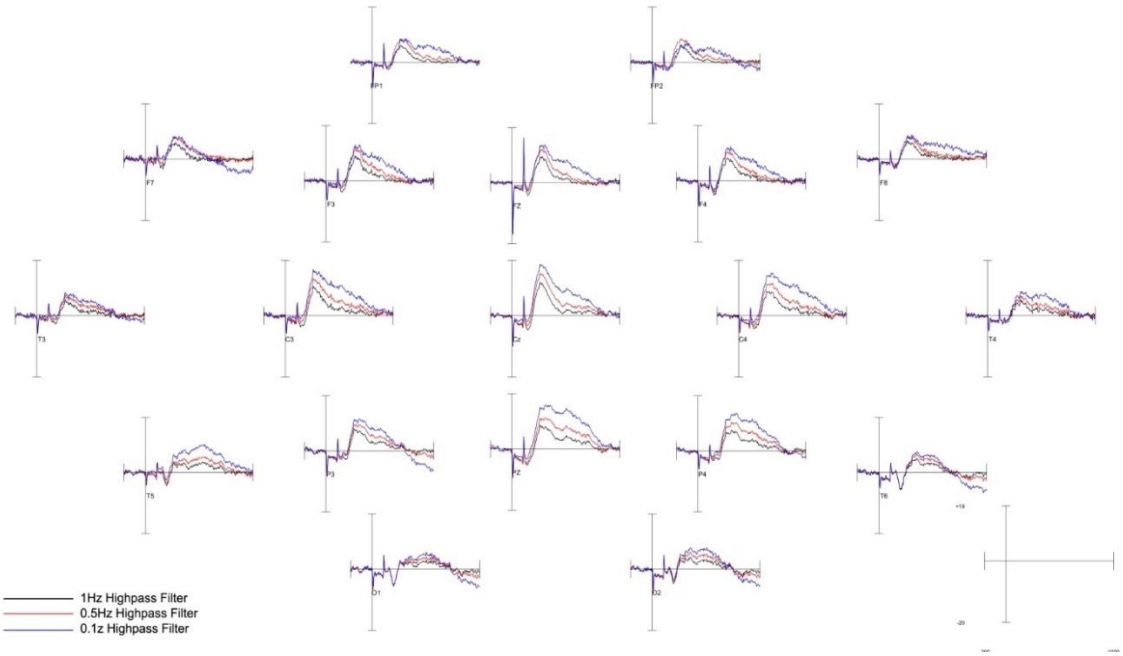
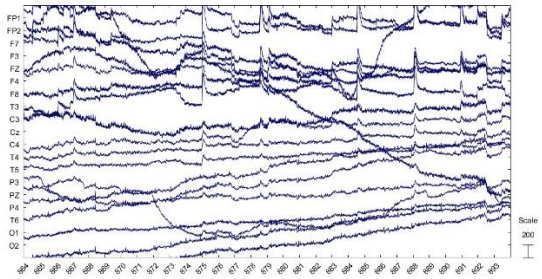


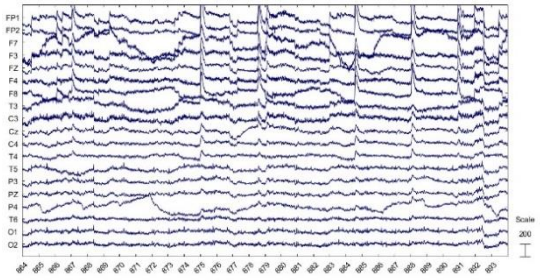
Figure 4.5: *Topographic representation of recorded electrodes for highpass filter comparisons when using a 50Hz Notch filter and ICA eye-based artifact rejection. A positive inflection starting at ~300ms when using the 0.1Hz highpass filter is seen in Fz, F4, F8, C3, Cz, C4, T4, T5 and Pz.*

Self-reported increased perspiration was captured as part of the SSQ administered after completion of the Chapter 3 arithmetic task. Of the 21 participants who completed the arithmetic task, 4 self-reported a slight increase to sweating, 2 self-reported a moderate increase, and 15 self-reported no sweating. Thirty second extracts from the continuous waveforms for no highpass filter, 0.1HzHp, 0.5HzHp and 1HzHp filtered data of two participants, one who reported no increase in sweating and one who reported a moderate increase in sweating are shown in Figures 4.6 and 4.7 respectively. Visual inspection of the continuous waveforms shows clear slow drift artifacts in unfiltered data in both participants (Figures 4.6a and 4.7a), which are modulated but not completely removed when using 0.1HzHp (Figures 4.6b and 4.7b). When A 0.5HzHp (Figures 4.6c and 4.7c) or 1HzHp (Figures 4.6d and 4.7d) is applied, the slow drift artifacts are no longer visible, with minor differences being visible between the higher frequency value highpass filters.

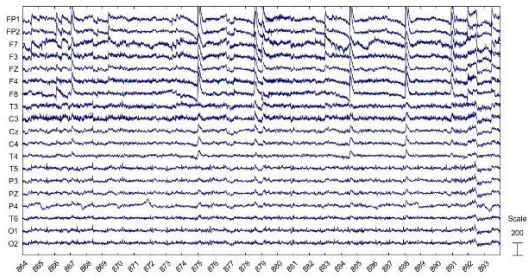
(a) No highpass filter



(b) 0.1Hz Highpass



(c) 0.5Hz Highpass



(d) 1Hz Highpass

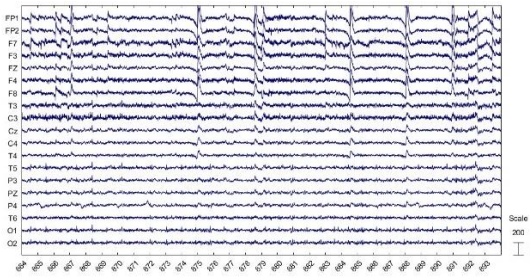
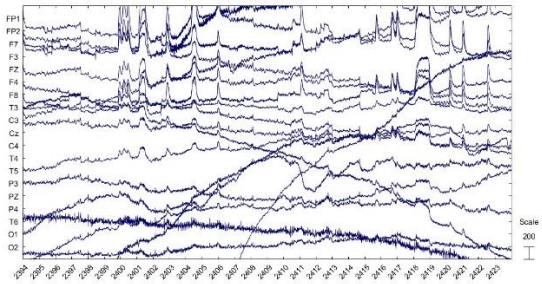
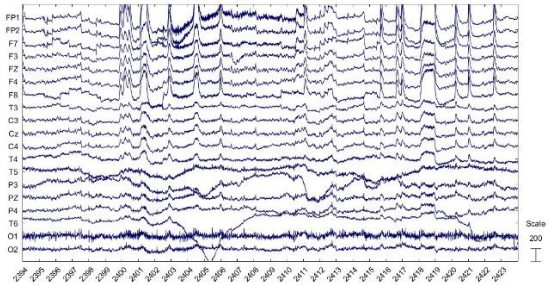


Figure 4.6: A comparison of different highpass filters on 30s of minimally processed data from a participant reporting no increase in sweating. The effect of no highpass filtering is seen in (a); a 0.1Hz highpass filter in (b); a 0.5Hz highpass filter in (c); and a 1Hz highpass filter in (d). Data is DC offset to see differences. Scale is set to 200mv.

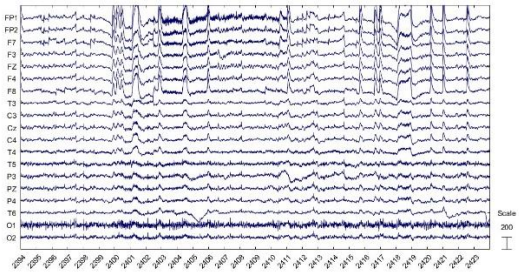
(a) No highpass filter



(b) 0.1Hz Highpass



(c) 0.5Hz Highpass



(d) 1Hz Highpass

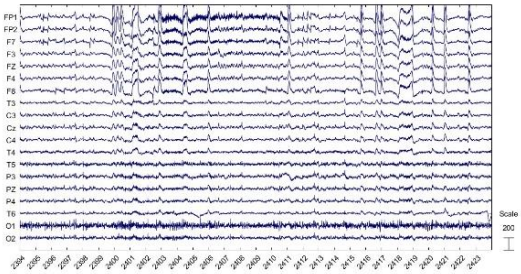


Figure 4.7: *Images of different highpass filters on 30s of minimally processed data of a participant reporting a moderate increase in sweating. The effect of no highpass filter is seen in figure (a); a 0.1Hz highpass filter in (b); a 0.5Hz highpass filter in (c); and a 1Hz highpass filter in (d). Data is DC offset to see differences. Scale is set to 200mv.*

4.3.2.3) Lowpass/Notch Filtering

Visual examination of the ERP waveforms from all recorded electrodes (Figure 4.8) shows that the 30Hz lowpass and 50Hz notch filters results in less high frequency contamination and thus a smoother waveform, but the ERPs otherwise closely overlap. Several electrodes including Fp2, F3, C3, C4, T5 and P4 show high levels of high frequency noise when NoLp/NF is applied, however this is largely attenuated when using the 50HzNF or removed when using the 30HzLp.

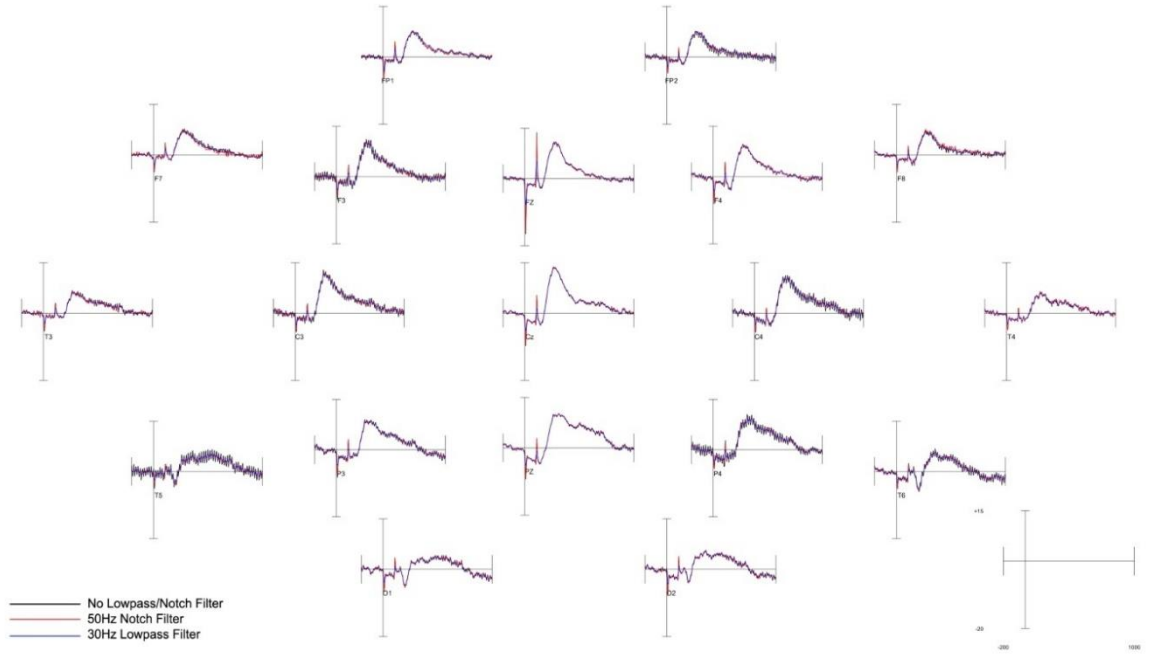


Figure 4.8: *Topographic representation of recorded electrodes for lowpass/notch comparisons when using a 0.5Hz highpass filter and ICA Eye-based artifact rejection. High-frequency noise can be identified in both hemispheres, and is primarily visible in the T5, P4, C4 and F3 electrodes.*

Inspection of the peaks in Figure 4.9 finds the 50Hz contamination is present when no lowpass/notch is applied (Figure 4.9a), modulated when applying the 50Hz notch filter (Figure 4.9b), and visibly present but functionally removed when using a 30HzLp filter (Figure 4.9c). The wider range of attenuated frequencies around and beyond 30HzLp is also visible in Figure 4.9c.

(a) No lowpass/notch filter (b) 50Hz Notch filter (c) 30Hz Lowpass filter

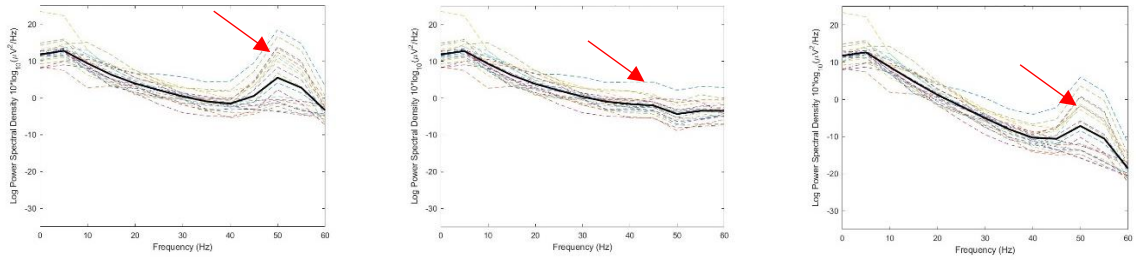


Figure 4.9: *Power spectral density graphs comparing 50Hz peaks between lowpass/notch filtering methods for 0.5Hz highpass filtered and ICA Eye-based artifact rejected data. The solid black line is the average of the individual datasets, which are represented by the surrounding dotted coloured lines. (a) No lowpass filter is applied; (b) 50Hz notch filter is applied; (c) 30Hz lowpass filter is applied. The 50Hz peaks are highlighted with red arrows.*

4.3.2.3) Eye Artifact Removal on the ERP Waveform

Visual inspection of Figure 4.12, which shows the effect of the eye-based artifacts captured in the current dataset across all highpass/lowpass combinations, finds an early negative inflection starting at ~120ms followed by a large positive inflection starting at 450ms and persisting until at least 1000ms when no EBAR has been removed. Moreover, the topographic array of the ERPs for each EBAR variation using 1HzHp and 30Hz lowpass filtering (Figure 4.10) shows this late positive inflection contaminated every electrode recorded. The late positive inflection largely disappears when eyeblinks were rejected, and has reduced amplitude towards the posterior of the head.

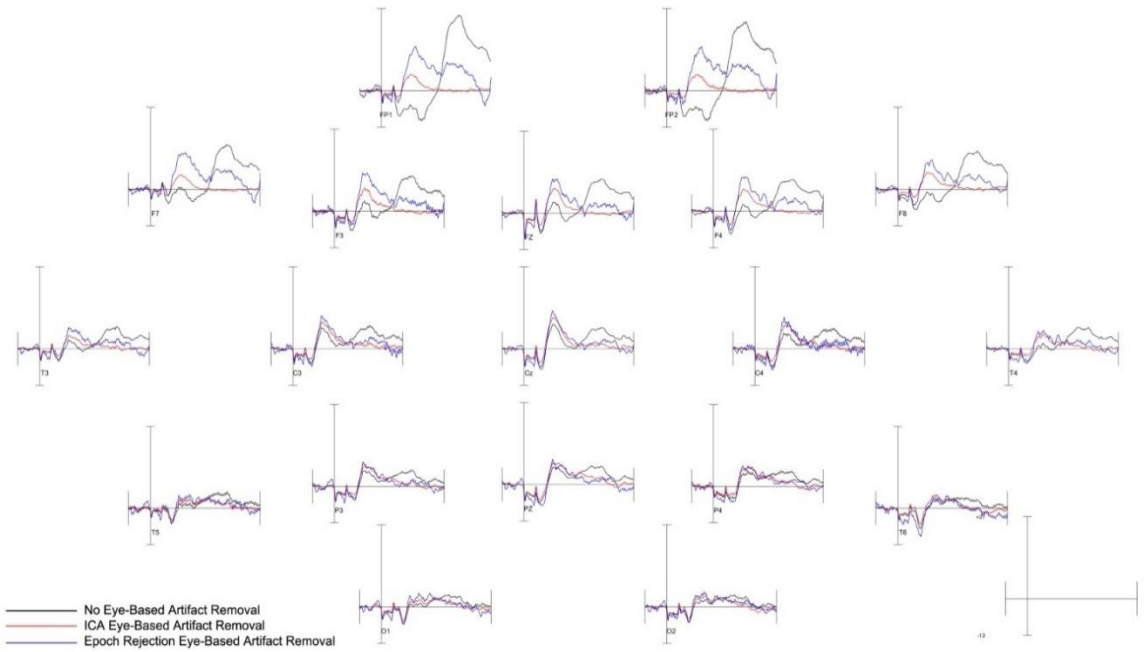


Figure 4.10: *Topographic representation of EBAR comparisons when using 1Hz highpass and a 30Hz lowpass filter. The filters were selected to remove any additional noise and isolate the effect of the eyeblink artifact on ERP recordings.*

The eyeblink component found comprises of an initial negative inflection covering both the N170 and P300 time ranges (seen clearest in the FP1 and FP2 electrodes in Figure 4.10), reducing the amplitude of the P300 component across the affected electrodes. This early negative inflection results in the smallest P300 peaks in all preprocessing pipeline variations for both the Fz and Cz electrodes (as seen in Figure 4.10), though not necessarily the Pz electrode where the power of the

eyeblink has lessened. Moreover, the negative inflection caused by the eyeblink artifact does not result in a significantly larger Cz N170 peak compared to Epoch-EBAR, but is significantly larger than the No-EBAR peak.

The effects of the ICA can be seen in Figure 4.11, which shows how eye-related artifacts were removed without removing the other features of the EEG waveform. The removal of eye-related components reduces the amplitude of the eyeblink artifacts, but the waveform remains largely overlapping with the un-adjusted data.

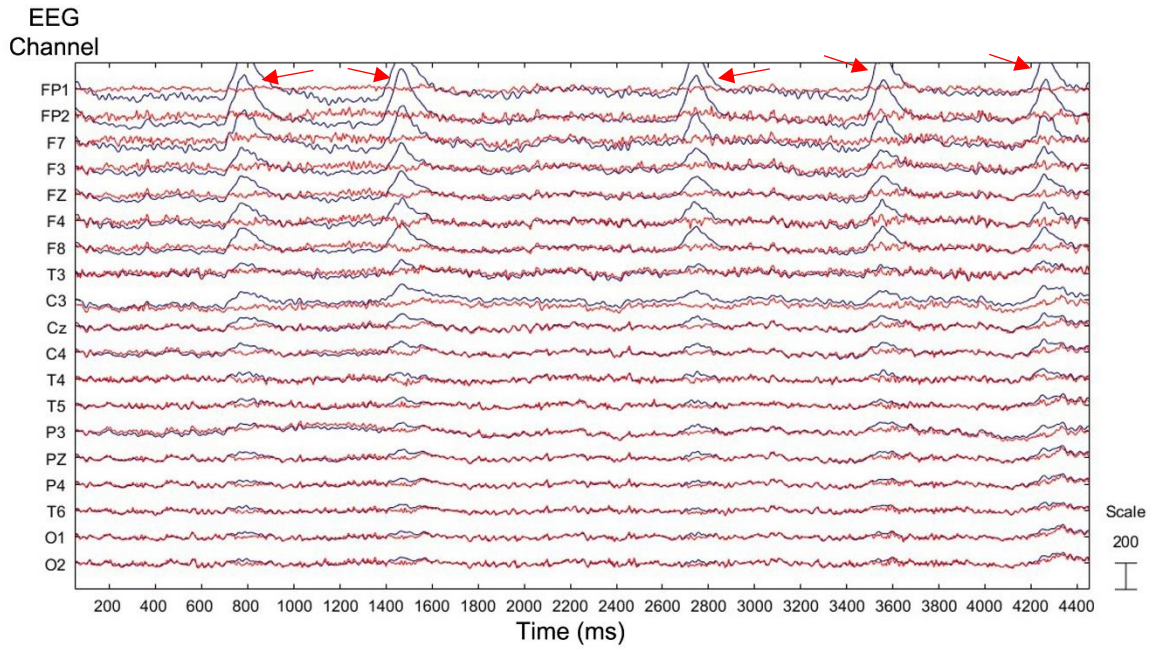


Figure 4.11: *Epoched data with (red line) and without (black line) eye-related components identified by ICA-EBAR removed. The data was filtered between 0.5-40Hz for clarity. Eye blinks visible when no ICA components have been removed are marked with red arrows.*

4.3.3) N170 and P300 Statistical Analysis Results

The grand average ERP waveforms of each preprocessing pipeline variant was generated for the Cz electrode, and organised to present the differences between highpass filter variations (Figure 4.12), lowpass/notch filter variations (Figure 4.13), and EBAR variations (Figure 4.14).

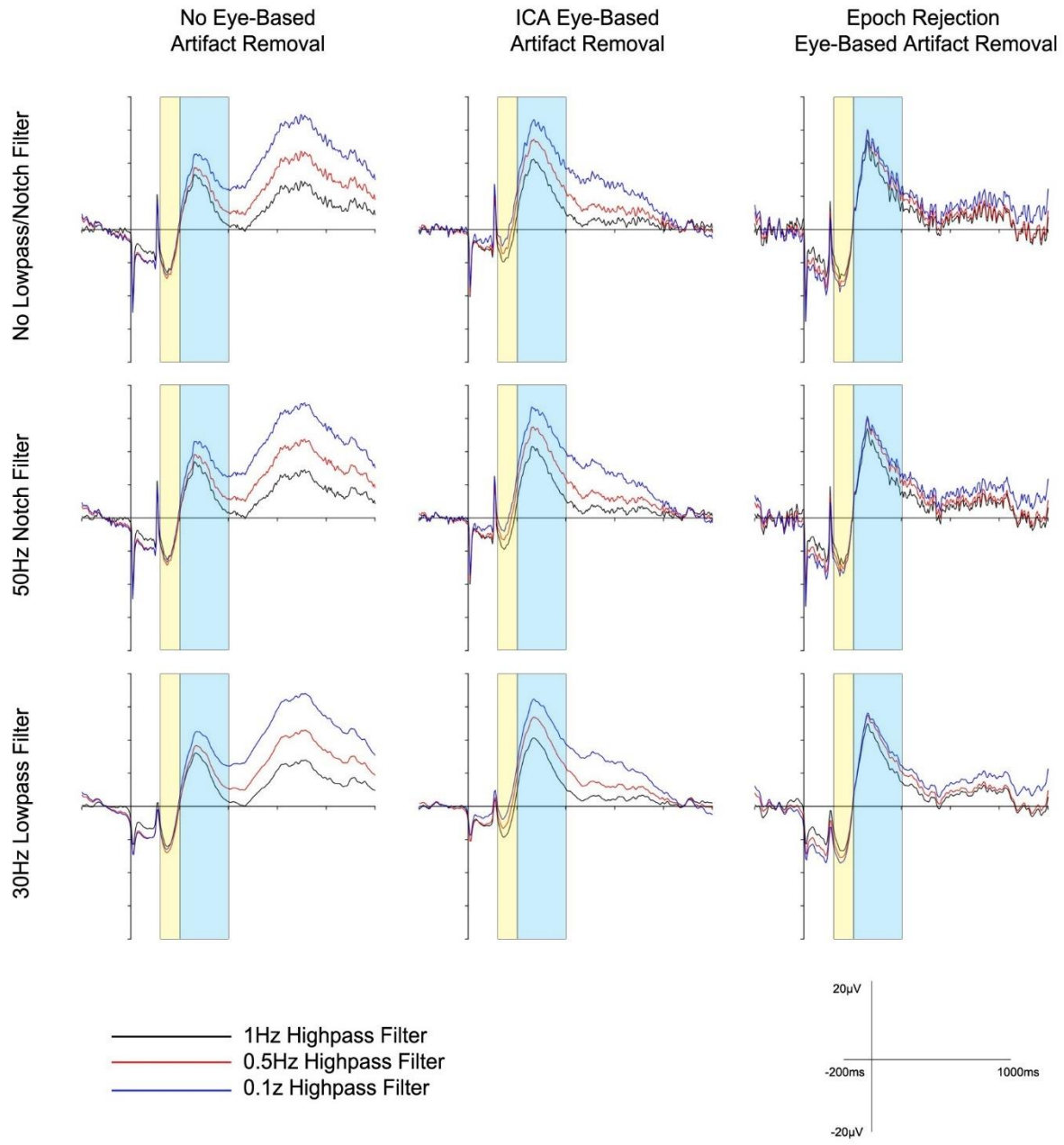


Figure 4.12: *Comparison between using a 0.1Hz, 0.5Hz and 1Hz highpass filter within each combination of lowpass/notch filtering and EBAR method for the Cz electrode. The highpass filter was applied before both the lowpass/notch filter and EBAR. The yellow time window is the N170 time range (120-200ms) and the cyan time window is the P300 time range (200-400ms). n=21.*

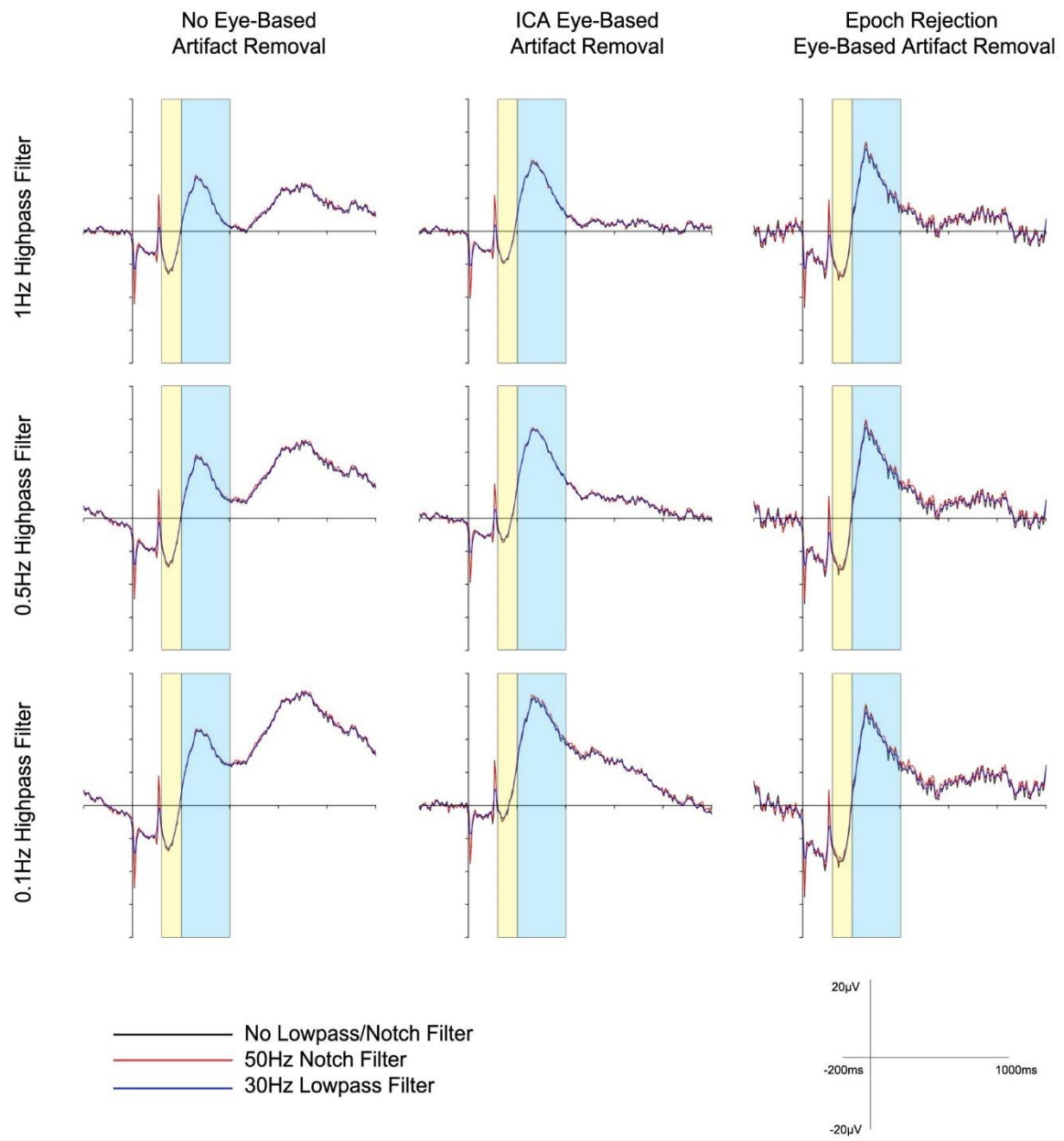


Figure 4.13: *Comparison between using a no lowpass or notch filter, a 50Hz notch filter and a 30Hz lowpass filter within each combination of highpass filtering and EBAR method for the Cz electrode. The highpass filter was applied after the lowpass/notch filter and before the EBAR. The yellow time window is the N170 time range (120-200ms) and the cyan time window is the P300 time range (200-400ms). n=21.*

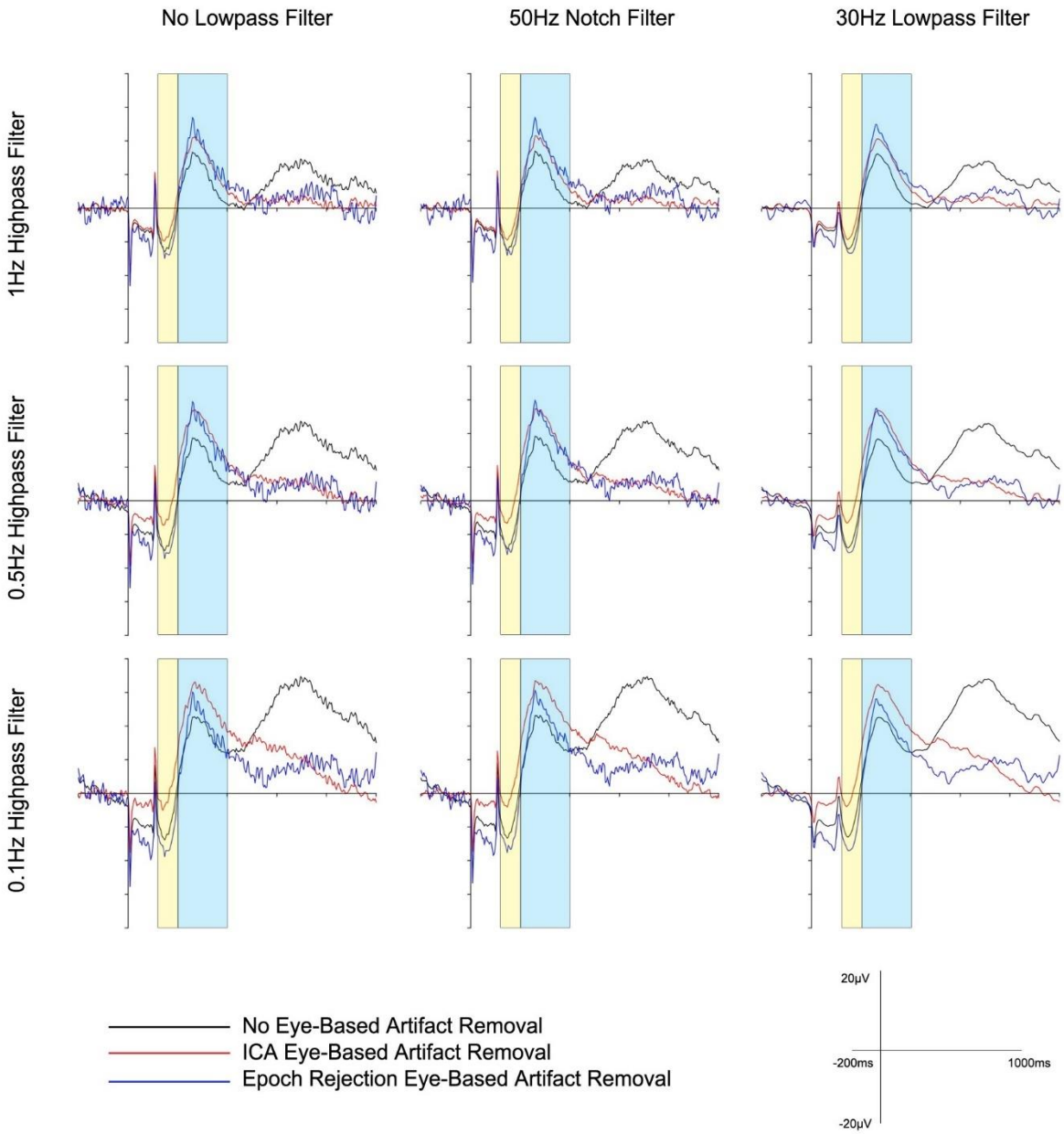


Figure 4.14: *Comparison between not removing eye-based artifacts, using ICA to remove eye-based artifacts, and rejecting contaminated epochs to remove eye-based artifacts within each combination of highpass filtering and lowpass/notch filter for the Cz electrode. The eye-based artifacts were removed after the highpass, and lowpass/notch filtering was performed. The yellow time window is the N170 time range (120-200ms), and the cyan time window is the P300 time range (200-400ms). n=21.*

4.3.3.1) N170 Peak Amplitude

For the N170 peak amplitude analysis within the Cz electrode, Mauchly's test of Sphericity is violated for all main effects and interactions: highpass ($\chi^2(2)=27.75$, $p\leq0.001$, $\varepsilon=0.57$); lowpass/notch ($\chi^2(2)=10.38$, $p=0.006$, $\varepsilon=0.7$); EBAR ($\chi^2(2)=17.98$, $p\leq0.001$, $\varepsilon=0.62$); highpass x lowpass/notch ($\chi^2(9)=57.29$, $p\leq0.001$, $\varepsilon=0.46$); highpass x EBAR ($\chi^2(9)=46.29$, $p\leq0.001$, $\varepsilon=0.47$); lowpass/notch x EBAR ($\chi^2(9)=111.47$, $p\leq0.001$, $\varepsilon=0.39$); highpass x lowpass/notch x EBAR ($\chi^2(35)=256.23$, $p\leq0.001$, $\varepsilon=0.29$). All violations were adjusted using the Greenhouse-Geisser correction. The significant main effects and interactions are summarised in table 4.1.

Table 4.1: *Summary of the main effects and interactions of the N170 peak amplitude ANOVA statistical analysis. n=21.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Highpass	0.17	1.13	22.63	0.71	0.01
Lowpass/Notch	35.44	1.41	28.15	**<0.001***	0.64
EBAR	11.34	1.24	24.82	**0.001***	0.36
Highpass x Lowpass/Notch	0.28	1.85	36.98	0.74	0.01
Highpass x EBAR	8.39	1.86	37.16	**0.001***	0.3
Lowpass/Notch x EBAR	6.52	1.57	31.44	**0.01***	0.25
Highpass x Lowpass/Notch x EBAR	0.41	2.28	45.69	0.69	0.02

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

The main effects of lowpass/notch and EBAR reached significance within the N170 peak amplitude ANOVA. Post hoc analysis within lowpass/notch comparisons found that using NoLp/NF (M=-10.1 ± 1.3) resulted in significantly larger negative peaks than 50HzNF (M=-8.9 ± 1.3) (t(20)=-1.148, p=0.003, d=-0.827) and 30HzLp (M=-7.6 ± 1.1) (t(20)=-2.446, p≤0.001, d=-1.492), with 30HzLp further having a smaller peak than 50HzNF (t(20)=-1.298, p≤0.001, d=-1.54). Between EBAR methods, it is found that using ICA-EBAR (M=-5.2 ± 0.7) resulted in significantly smaller peaks than both

No-EBAR (M=-8.4 ± 1.1) (t(20)=-3.213, p=0.002, d=-0.889) and Epoch-EBAR (M=-13.1 ± 2.3) (t(20)=7.864, p=0.003, d=0.849).

The interaction between highpass x EBAR was significant. Post hoc comparisons (table 4.2) found within EBAR and between highpass, peak amplitude significantly differed between No-EBAR 1HzHp & No-EBAR 0.5HzHp, and between ICA-EBAR 1HzHp with both ICA-EBAR 0.5HzHp & ICA-EBAR 0.1HzHp. Within highpass and between EBAR methods, it is found that 1HzHp ICA-EBAR resulted in a significantly smaller peak than 1HzHp Epoch-EBAR, that 0.5HzHp ICA-EBAR (M=-5.1 ± 0.7) resulted in a significantly smaller peak than both 0.5HzHp No-EBAR & 0.5HzHp Epoch-EBAR, and that 0.1HzHp ICA-EBAR had a significantly smaller peak than 0.1HzHp No-EBAR and 0.1HzHp Epoch-EBAR.

Table 4.2: *Post hoc comparisons within the significant interaction of EBAR and Highpass filter within the N170 peak amplitude ANOVA analysis. n=21.*

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
EBAR Highpass	No-EBAR 1HzHp	-7.95	0.97	No-EBAR 0.5HzHp	-8.87	1.21	2.74	20	0.92	0.34	**0.04***
	No-EBAR 1HzHp	-7.95	0.97	No-EBAR 0.1HzHp	-8.41	1.36	0.6	20	0.46	0.77	1
	No-EBAR 0.5HzHp	-8.87	1.21	No-EBAR 0.1HzHp	-8.41	1.36	-0.83	20	-0.46	0.56	1
	ICA-EBAR 1HzHp	-6.51	0.67	ICA-EBAR 0.5HzHp	-5.14	0.66	-3.74	20	-1.38	0.37	**0.004***
	ICA-EBAR 1HzHp	-6.51	0.67	ICA-EBAR 0.1HzHp	-3.94	0.96	-3.44	20	-2.58	0.75	**0.01***
	ICA-EBAR 0.5HzHp	-5.14	0.66	ICA-EBAR 0.1HzHp	-3.94	0.96	-1.83	20	-1.2	0.66	0.25
	Epoch-EBAR 1HzHp	-12.18	2.12	Epoch-EBAR 0.5HzHp	-13.14	2.36	1.79	20	0.97	0.54	0.26
	Epoch-EBAR 1HzHp	-12.18	2.12	Epoch-EBAR 0.1HzHp	-13.86	2.69	1.41	20	1.68	1.2	0.52
	Epoch-EBAR 0.5HzHp	-13.14	2.36	Epoch-EBAR 0.1HzHp	-13.86	2.69	0.96	20	0.71	0.74	1
Highpass EBAR	1HzHp No-EBAR	-7.95	0.97	1HzHp ICA-EBAR	-6.51	0.67	-2.23	20	-1.43	0.64	0.11
	1HzHp No-EBAR	-7.95	0.97	1HzHp Epoch-EBAR	-12.18	2.12	2.49	20	4.23	1.7	0.07
	1HzHp ICA-EBAR	-6.51	0.67	1HzHp Epoch-EBAR	-12.18	2.12	3.38	20	5.66	1.67	**0.01***
	0.5HzHp No-EBAR	-8.87	1.21	0.5HzHp ICA-EBAR	-5.14	0.66	-3.99	20	-3.73	0.94	**0.002***
	0.5HzHp No-EBAR	-8.87	1.21	0.5HzHp Epoch-EBAR	-13.14	2.36	2.23	20	4.27	1.92	0.11
	0.5HzHp ICA-EBAR	-5.14	0.66	0.5HzHp Epoch-EBAR	-13.14	2.36	3.78	20	8.01	2.12	**0.003***
	0.1HzHp No-EBAR	-8.41	1.36	0.1HzHp ICA-EBAR	-3.94	0.96	-4.11	20	-4.47	1.09	**0.002***
	0.1HzHp No-EBAR	-8.41	1.36	0.1HzHp Epoch-EBAR	-13.86	2.69	2.59	20	5.45	2.11	0.053
	0.1HzHp ICA-EBAR	-3.94	0.96	0.1HzHp Epoch-EBAR	-13.86	2.69	4.1	20	9.92	2.42	**0.002***

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, p<0.05.

The interaction between lowpass/notch x EBAR also reached significance, summarised in table 4.3. The main findings are that, within EBAR method and between lowpass/notch filters, all interactions reached significance, with NoLp/NF

consistently resulting in the largest peak and 30HzLp resulting in the smallest peak amplitudes. Within lowpass/notch and between EBAR, it is found that ICA-EBAR resulted in significantly smaller peaks compared to No-EBAR and Epoch-EBAR across all lowpass/notch variants. The comparison between No-EBAR and Epoch-EBAR reaches significance in NoLp/NF only, with NoLp/NF No-EBAR (M=-9.2 ± 1.1) resulting in a smaller peak than NoLp/NF Epoch-EBAR (M=-15.1 ± 2.7) ($t(20)=5.87$, $p=0.048$, $d=0.575$).

Table 4.3: *Post hoc comparisons within the significant interaction of EBAR and lowpass/notch filter within the N170 peak amplitude ANOVA analysis. n=21.*

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
EBAR Lowpass/ Notch	No-EBAR NoLp/NF	-9.2	1.12	No-EBAR 50HzNF	-8.5	1.2	-2.93	20	-0.7	0.24	**0.03***
	No-EBAR NoLp/NF	-9.2	1.12	No-EBAR 30HzLp	-7.52	1.13	-8.01	20	-1.68	0.21	**<0.001***
	No-EBAR 50HzNF	-8.5	1.2	No-EBAR 30HzLp	-7.52	1.13	-6.74	20	-0.98	0.15	**<0.001***
	ICA-EBAR NoLp/NF	-5.98	0.67	ICA-EBAR 50HzNF	-5.24	0.75	-3.03	20	-0.75	0.25	**0.02***
	ICA-EBAR NoLp/NF	-5.98	0.67	ICA-EBAR 30HzLp	-4.36	0.69	-7.42	20	-1.62	0.22	**<0.001***
	ICA-EBAR 50HzNF	-5.24	0.75	ICA-EBAR 30HzLp	-4.36	0.69	-6.44	20	-0.87	0.14	**<0.001***
	Epoch-EBAR NoLp/NF	-15.07	2.69	Epoch-EBAR 50HzNF	-13.07	2.33	-2.97	20	-2	0.67	**0.02***
	Epoch-EBAR NoLp/NF	-15.07	2.69	Epoch-EBAR 30HzLp	-11.03	2.09	-4.74	20	-4.04	0.85	**<0.001***
	Epoch-EBAR 50HzNF	-13.07	2.33	Epoch-EBAR 30HzLp	-11.03	2.09	-4.43	20	-2.04	0.46	**0.004***
Lowpass/ Notch EBAR	NoLp/NF No-EBAR	-9.2	1.12	NoLp/NF ICA-EBAR	-5.98	0.67	-4.14	20	-3.22	0.78	**0.002***
	NoLp/NF No-EBAR	-9.2	1.12	NoLp/NF Epoch-EBAR	-15.07	2.69	2.64	20	5.87	2.23	**0.048***
	NoLp/NF ICA-EBAR	-5.98	0.67	NoLp/NF Epoch-EBAR	-15.07	2.69	3.86	20	9.09	2.35	**0.003***
	50HzNF No-EBAR	-8.5	1.2	50HzNF ICA-EBAR	-5.24	0.75	-3.99	20	-3.27	0.82	**0.002***
	50HzNF No-EBAR	-8.5	1.2	50HzNF Epoch-EBAR	-13.07	2.33	2.44	20	4.57	1.87	0.07
	50HzNF ICA-EBAR	-5.24	0.75	50HzNF Epoch-EBAR	-13.07	2.33	3.95	20	7.83	1.98	**0.002***
	30HzLp No-EBAR	-7.52	1.13	30HzLp ICA-EBAR	-4.36	0.69	-4.08	20	-3.16	0.78	**0.002***
	30HzLp No-EBAR	-7.52	1.13	30HzLp Epoch-EBAR	-11.03	2.09	2.15	20	3.52	1.63	0.13
	30HzLp ICA-EBAR	-4.36	0.69	30HzLp Epoch-EBAR	-11.03	2.09	3.73	20	6.67	1.79	**0.004***

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, p<0.05.

4.3.3.2) N170 Mean Amplitude

Within the N170 mean amplitude analysis, the assumption of sphericity is violated for all main effects and interactions: highpass ($\chi^2(2)=24.185$, $p\leq0.001$, $\varepsilon=0.581$); lowpass/notch ($\chi^2(2)=15.908$, $p\leq0.001$, $\varepsilon=0.638$); EBAR ($\chi^2(2)=9.597$, $p=0.008$, $\varepsilon=0.716$); highpass x lowpass/notch ($\chi^2(9)=60.271$, $p\leq0.001$, $\varepsilon=0.367$); highpass x EBAR ($\chi^2(9)=43.769$, $p\leq0.001$, $\varepsilon=0.468$); lowpass/notch x EBAR ($\chi^2(9)=45.153$, $p\leq0.001$, $\varepsilon=0.557$); highpass x lowpass/notch x EBAR ($\chi^2(35)=438.664$, $p\leq0.001$, $\varepsilon=0.292$). All violations were adjusted using the Greenhouse-Geisser correction. The significant main effects and interactions are summarised in Table 4.4.

Table 4.4: *Summary of the main effects and interactions of the N170 mean amplitude analysis. n=21.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Highpass	0.62	1.16	23.26	0.46	0.03
Lowpass/Notch	16.11	1.28	25.53	**<0.001***	0.45
EBAR	7.32	1.43	28.64	**0.01***	0.27
Highpass X Lowpass/Notch	3.74	1.47	29.33	**0.048***	0.16
Highpass x EBAR	10.67	1.87	37.42	**<0.001***	0.35
Lowpass/Notch x EBAR	4.48	2.23	44.58	**0.01***	0.18
Highpass x Lowpass/Notch x EBAR	0.94	2.34	46.74	0.41	0.05

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

The main effects of lowpass/notch and EBAR reach significance in the N170 mean amplitude analysis. Post hoc comparisons find that between lowpass/notch, using NoLp/NF (M=-3.9 ± 1) results in significantly larger mean amplitudes than 50HzNF (M=-3.6 ± 1) and 30HzLp (M=-3.6 ± 1) filters. Moreover, between EBAR methods, it is found that using ICA-EBAR (M=-1 ± 0.7) results in significantly smaller mean amplitudes than both No-EBAR (M=-4.3 ± 1.1) ($t(20)=-3.321$, $p=0.001$, $d=-0.938$) and Epoch-EBAR (M=-5.7 ± 1.7) ($t(20)=4.666$, $p=0.015$, $d=0.685$).

Several interactions reached significance within the N170 mean amplitude analysis. The significant interaction between highpass x lowpass/notch is summarised in Table 4.5, however no significant post hoc comparisons are found within lowpass/notch and between highpass. The main findings within highpass and between lowpass/notch is that each comparison utilising NoLp reaches significance for post hoc comparisons, resulting in larger mean amplitudes than both 50HzNF and 30HzLp across all highpass filtering variations and subsequently matching the results found in the main effect of lowpass/notch.

Table 4.5: *Post hoc comparisons within the significant interaction of lowpass/notch filter and highpass filter within the N170 mean amplitude ANOVA analysis. n=21.*

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
Lowpass/Notch Highpass	NoLP/NF 1HzHp	-4.03	0.86	NoLP/NF 0.5HzHp	-4.02	1.01	-0.03	20	-0.01	0.3	1
	NoLP/NF 1HzHp	-4.03	0.86	NoLP/NF 0.1HzHp	-3.54	1.27	-0.7	20	-0.5	0.71	1
	NoLP/NF 0.5HzHp	-4.02	1.01	NoLP/NF 0.1HzHp	-3.54	1.27	-0.97	20	-0.49	0.5	1
	50HzNF 1HzHp	-3.81	0.88	50HzNF 0.5HzHp	-3.71	1.04	-0.34	20	-0.11	0.31	1
	50HzNF 1HzHp	-3.81	0.88	50HzNF 0.1HzHp	-3.22	1.29	-0.85	20	-0.6	0.7	1
	50HzNF 0.5HzHp	-3.71	1.04	50HzNF 0.1HzHp	-3.22	1.29	-0.99	20	-0.49	0.49	1
	30HzLp 1HzHp	-3.79	0.84	30HzLp 0.5HzHp	-3.79	1	-0.01	20	0	0.3	1
	30HzLp 1HzHp	-3.79	0.84	30HzLp 0.1HzHp	-3.3	1.26	-0.69	20	-0.49	0.71	1
	30HzLp 0.5HzHp	-3.79	1	30HzLp 0.1HzHp	-3.3	1.26	-0.98	20	-0.49	0.5	1
Highpass Lowpass/Notch	1HzHp NoLP/NF	-4.03	0.86	1HzHp 50HzNF	-3.81	0.88	-5.28	20	-0.22	0.04	**<0.001***
	1HzHp NoLP/NF	-4.03	0.86	1HzHp 30HzLp	-3.79	0.84	-7.12	20	-0.24	0.03	**<0.001***
	1HzHp 50HzNF	-3.81	0.88	1HzHp 30HzLp	-3.79	0.84	-0.34	20	-0.02	0.05	1
	0.5HzHp NoLP/NF	-4.02	1.01	0.5HzHp 50HzNF	-3.71	1.04	-5.65	20	-0.32	0.06	**<0.001***
	0.5HzHp NoLP/NF	-4.02	1.01	0.5HzHp 30HzLp	-3.79	1	-9.21	20	-0.23	0.03	**<0.001***
	0.5HzHp 50HzNF	-3.71	1.04	0.5HzHp 30HzLp	-3.79	1	1.32	20	0.09	0.06	0.61
	0.1HzHp NoLP/NF	-3.54	1.27	0.1HzHp 50HzNF	-3.22	1.29	-3.93	20	-0.32	0.08	**0.002***
	0.1HzHp NoLP/NF	-3.54	1.27	0.1HzHp 30HzLp	-3.3	1.26	-7.54	20	-0.24	0.03	**<0.001***
	0.1HzHp 50HzNF	-3.22	1.29	0.1HzHp 30HzLp	-3.3	1.26	0.9	20	0.08	0.09	1

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, $p<0.05$.

A second significant interaction between highpass x EBAR (table 4.6) is found. Post hoc comparisons within EBAR and between highpass finds highpass filtering only

differs within ICA-EBAR, with ICA-EBAR 1HzHp resulting in a significantly larger mean amplitude than both ICA-EBAR 0.5HzHp and ICA-EBAR 0.1HzHp. Moreover, ICA-EBAR 0.5HzHp produces a significantly larger mean amplitude than ICA-EBAR 0.1HzHp. Within highpass and within EBAR, it is found ICA-EBAR results in significantly smaller mean amplitude than both No-EBAR and Epoch-EBAR within both 0- 5HzHp and 0.1Hz, but not within 1HzHp filtering: 0.5HzHp ICA-EBAR < 0.5HzHp No-EBAR & 0.5HzHp Epoch-EBAR; 0.1HzHp ICA-EBAR < 0.1HzHp No-EBAR & 0.1HzHp Epoch-EBAR.

Table 4.6: *Post hoc comparisons within the significant interaction of EBAR and highpass filter within the N170 mean amplitude ANOVA analysis. n=21.*

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
EBAR Highpass	No-EBAR 1HzHp	-4.1	0.9	No-EBAR 0.5HzHp	-4.8	1.18	1.96	20	0.7	0.36	0.19
	No-EBAR 1HzHp	-4.1	0.9	No-EBAR 0.1HzHp	-4.15	1.38	0.06	20	0.05	0.83	1
	No-EBAR 0.5HzHp	-4.8	1.18	No-EBAR 0.1HzHp	-4.15	1.38	-1.07	20	-0.65	0.61	0.89
	ICA-EBAR 1HzHp	-2.62	0.64	ICA-EBAR 0.5HzHp	-0.98	0.69	-4.02	20	-1.64	0.41	**0.002***
	ICA-EBAR 1HzHp	-2.62	0.64	ICA-EBAR 0.1HzHp	0.52	0.95	-4.83	20	-3.14	0.65	**<0.001***
	ICA-EBAR 0.5HzHp	-0.98	0.69	ICA-EBAR 0.1HzHp	0.52	0.95	-2.7	20	-1.5	0.56	**0.04***
	Epoch-EBAR 1HzHp	-4.92	1.39	Epoch-EBAR 0.5HzHp	-5.74	1.75	1.58	20	0.82	0.52	0.39
	Epoch-EBAR 1HzHp	-4.92	1.39	Epoch-EBAR 0.1HzHp	-6.42	2.14	1.32	20	1.5	1.14	0.61
	Epoch-EBAR 0.5HzHp	-5.74	1.75	Epoch-EBAR 0.1HzHp	-6.42	2.14	0.95	20	0.68	0.72	1
Highpass EBAR	1HzHp No-EBAR	-4.1	0.9	1HzHp ICA-EBAR	-2.62	0.64	-2.48	20	-1.48	0.6	0.07
	1HzHp No-EBAR	-4.1	0.9	1HzHp Epoch-EBAR	-4.92	1.39	0.75	20	0.82	1.1	1
	1HzHp ICA-EBAR	-2.62	0.64	1HzHp Epoch-EBAR	-4.92	1.39	2.09	20	2.3	1.1	0.15
	0.5HzHp No-EBAR	-4.8	1.18	0.5HzHp ICA-EBAR	-0.98	0.69	-4.14	20	-3.82	0.92	**0.002***
	0.5HzHp No-EBAR	-4.8	1.18	0.5HzHp Epoch-EBAR	-5.74	1.75	0.64	20	0.94	1.47	1
	0.5HzHp ICA-EBAR	-0.98	0.69	0.5HzHp Epoch-EBAR	-5.74	1.75	3.05	20	4.76	1.56	**0.02***
	0.1HzHp No-EBAR	-4.15	1.38	0.1HzHp ICA-EBAR	0.52	0.95	-4.35	20	-4.67	1.07	**0.001***
	0.1HzHp No-EBAR	-4.15	1.38	0.1HzHp Epoch-EBAR	-6.42	2.14	1.39	20	2.27	1.63	0.54
	0.1HzHp ICA-EBAR	0.52	0.95	0.1HzHp Epoch-EBAR	-6.42	2.14	3.56	20	6.94	1.95	**0.01***

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, p<0.05.

The interaction between lowpass/notch x EBAR reaches significance, and is summarised in table 4.7. The main findings in the interaction between lowpass/notch x EBAR is that with between lowpass/notch, NoLp/NF results in significantly larger

mean amplitude than both 50HzNF and 30HzLp across EBAR methods, mirroring the comparisons within the main effect of lowpass/notch filters. Moreover, within lowpass/notch and between EBAR it is found that using ICA-EBAR results in a lower mean amplitude than both No-EBAR and Epoch-EBAR across all lowpass/notch filtering methods, matching the comparisons within the main effect of EBAR.

Table 4.7: *Post hoc comparisons within the significant interaction of EBAR and lowpass/notch filter within the N170 mean amplitude ANOVA analysis. n=21.*

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
EBAR Lowpass/Notch	No-EBAR NoLP/NF	-4.5	1.11	No-EBAR 50HzNF	-4.28	1.14	-2.64	20	-0.22	0.09	**0.047***
	No-EBAR NoLP/NF	-4.5	1.11	No-EBAR 30HzLp	-4.27	1.1	-9.08	20	-0.23	0.03	**<0.001***
	No-EBAR 50HzNF	-4.28	1.14	No-EBAR 30HzLp	-4.27	1.1	-0.12	20	-0.01	0.09	1
	ICA-EBAR NoLP/NF	-1.16	0.7	ICA-EBAR 50HzNF	-0.95	0.72	-3.82	20	-0.2	0.05	**0.003***
	ICA-EBAR NoLP/NF	-1.16	0.7	ICA-EBAR 30HzLp	-0.97	0.69	-5.96	20	-0.18	0.03	**<0.001***
	ICA-EBAR 50HzNF	-0.95	0.72	ICA-EBAR 30HzLp	-0.97	0.69	0.32	20	0.02	0.06	1
	Epoch-EBAR NoLP/NF	-5.93	1.72	Epoch-EBAR 50HzNF	-5.5	1.74	-5.78	20	-0.43	0.07	**<0.001***
	Epoch-EBAR NoLP/NF	-5.93	1.72	Epoch-EBAR 30HzLp	-5.64	1.7	-7.33	20	-0.29	0.04	**<0.001***
	Epoch-EBAR 50HzNF	-5.5	1.74	Epoch-EBAR 30HzLp	-5.64	1.7	1.81	20	0.14	0.08	0.25
Lowpass/Notch EBAR	NoLP/NF No-EBAR	-4.5	1.11	NoLP/NF ICA-EBAR	-1.16	0.7	-4.33	20	-3.34	0.77	**0.001***
	NoLP/NF No-EBAR	-4.5	1.11	NoLP/NF Epoch-EBAR	-5.93	1.72	1.04	20	1.43	1.38	0.94
	NoLP/NF ICA-EBAR	-1.16	0.7	NoLP/NF Epoch-EBAR	-5.93	1.72	3.22	20	4.78	1.48	**0.01***
	50HzNF No-EBAR	-4.28	1.14	50HzNF ICA-EBAR	-0.95	0.72	-4.25	20	-3.33	0.78	**0.001***
	50HzNF No-EBAR	-4.28	1.14	50HzNF Epoch-EBAR	-5.5	1.74	0.88	20	1.23	1.4	1
	50HzNF ICA-EBAR	-0.95	0.72	50HzNF Epoch-EBAR	-5.5	1.74	3.05	20	4.55	1.49	**0.02***
	30HzLp No-EBAR	-4.27	1.1	30HzLp ICA-EBAR	-0.97	0.69	-4.3	20	-3.29	0.77	**0.001***
	30HzLp No-EBAR	-4.27	1.1	30HzLp Epoch-EBAR	-5.64	1.7	1	20	1.38	1.38	0.99
	30HzLp ICA-EBAR	-0.97	0.69	30HzLp Epoch-EBAR	-5.64	1.7	3.15	20	4.67	1.48	**0.02***

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, p<0.05.

4.3.3.3) P300 Peak Amplitude

Within the P300 peak amplitude analysis, Mauchly's test of Sphericity is violated for all main effects and interactions: highpass ($\chi^2(2)=28.205$, $p \leq 0.001$, $\varepsilon=0.564$);

lowpass/notch ($\chi^2(2)=21.985$, $p \leq 0.001$, $\varepsilon=0.593$); EBAR ($\chi^2(2)=7.693$, $p=0.021$, $\varepsilon=0.75$); highpass x lowpass/notch ($\chi^2(9)=53.803$, $p \leq 0.001$, $\varepsilon=0.503$); highpass x EBAR ($\chi^2(9)=43.492$, $p \leq 0.001$, $\varepsilon=0.502$); lowpass/notch x EBAR ($\chi^2(9)=69.709$, $p \leq 0.001$, $\varepsilon=0.386$); highpass x lowpass/notch x EBAR ($\chi^2(35)=152.588$, $p \leq 0.001$, $\varepsilon=0.331$). The main effect of EBAR is adjusted using the Huynh-Feldt correction, whilst the remaining violations were adjusted with the Greenhouse-Geisser correction. The significant main effects and interactions are summarised in table 4.8.

Table 4.8: *Summary of the main effects and interactions of the P300 peak amplitude analysis. n=21.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Highpass	10.34	1.13	22.56	**0.003***	0.34
Lowpass/Notch	33.23	1.19	23.73	**<0.001***	0.62
EBAR	7.74	1.6	31.9	**0.003***	0.28
Highpass x Lowpass/Notch	1.41	2.01	40.24	0.26	0.07
Highpass x EBAR	4.51	2.01	40.17	**0.02***	0.18
Lowpass/Notch x EBAR	5.72	1.54	30.85	**0.01***	0.22
Highpass x Lowpass/Notch x EBAR	1.19	2.65	53.03	0.32	0.06

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, $p<0.05$.

Each main effect reached significance in the P300 peak amplitude analysis: highpass; lowpass/notch; and EBAR. Post hoc comparisons between highpass filters found that using a 1HzHp (M=14.3 ± 0.9) filter resulted in significantly smaller peaks than both 0.5HzHp (M=16.4 ± 1.1) ($t(20)=-2.094$, $p \leq 0.001$, $d=-1.095$) and 0.1HzHp (M=18.6 ± 1.6) ($t(20)=-4.292$, $p=0.007$, $d=-0.759$). Between lowpass/notch filters, it is found that using the 30HzLp (M=15.1 ± 1.1) resulted in significantly smaller peaks than both NoLp/NF (M=17.5 ± 1.2) ($t(20)=2.456$, $p \leq 0.001$, $d=1.488$) and 50HzNF (M=16.7 ± 1.1) ($t(20)=1.653$, $p \leq 0.001$, $d=2.821$). Between EBAR methods, post hoc comparisons find using No-EBAR (M=13.5 ± 1.1) results in significantly reduced

peaks compared to both ICA-EBAR (M=17.1 ± 1.1) (t(20)=-3.624, p=0.006, d=-0.779) and Epoch-EBAR (M=18.8 ± 1.8) (t(20)=-5.318, p=0.017, d=-0.679).

Two interactions reach significance within the P300 peak amplitude analysis. The first interaction to reach significance is between highpass x EBAR, summarised in table 4.9. Post hoc comparisons within EBAR and between highpass find that using No-EBAR 1HzHp results in significantly smaller peaks than No-EBAR 0.5HzHp and No-EBAR 0.1HzHp; using ICA-EBAR results in significantly smaller peaks than both ICA-EBAR 0.5HzHp and ICA-EBAR 0.1HzHp; and using Epoch-EBAR 1HzHp results in smaller peaks than Epoch-EBAR 0.5HzHp only. Within highpass and between EBAR, significant post hoc comparisons include 1HzHp Epoch-EBAR resulting in a larger peak than 1HzHp No-EBAR and 1HzHp ICA-EBAR, 0.5HzHp No-EBAR resulting in a smaller peak than both 0.5HzHp ICA-EBAR and 0.5HzHp Epoch-EBAR; and 0.1HzHp ICA-EBAR resulting in a larger peak than 0.1HzHp No-EBAR.

Table 4.9: Post hoc comparisons within the significant interaction of EBAR and Highpass filter within the P300 peak amplitude ANOVA analysis. n=21

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
EBAR Highpass	No-EBAR 1HzHp	11.59	1.05	No-EBAR 0.5HzHp	13.08	1.13	-2.8	20	-1.5	0.54	0.03*
	No-EBAR 1HzHp	11.59	1.05	No-EBAR 0.1HzHp	15.76	1.54	-2.76	20	-4.17	1.51	0.04*
	No-EBAR 0.5HzHp	13.08	1.13	No-EBAR 0.1HzHp	15.76	1.54	-2.45	20	-2.67	1.09	0.07
	ICA-EBAR 1HzHp	13.9	0.99	ICA-EBAR 0.5HzHp	17.06	1.22	-6.31	20	-3.16	0.5	<0.001*
	ICA-EBAR 1HzHp	13.9	0.99	ICA-EBAR 0.1HzHp	20.33	1.57	-5.39	20	-6.42	1.19	<0.001*
	ICA-EBAR 0.5HzHp	17.06	1.22	ICA-EBAR 0.1HzHp	20.33	1.57	-2.48	20	-3.26	1.32	0.07
	Epoch-EBAR 1HzHp	17.49	1.34	Epoch-EBAR 0.5HzHp	19.12	1.75	-2.64	20	-1.63	0.61	0.047*
	Epoch-EBAR 1HzHp	17.49	1.34	Epoch-EBAR 0.1HzHp	19.77	2.49	-1.45	20	-2.28	1.57	0.48
	Epoch-EBAR 0.5HzHp	19.12	1.75	Epoch-EBAR 0.1HzHp	19.77	2.49	-0.62	20	-0.66	1.06	1
Highpass EBAR	1HzHp No-EBAR	11.59	1.05	1HzHp ICA-EBAR	13.9	0.99	-2.53	20	-2.32	0.92	0.06
	1HzHp No-EBAR	11.59	1.05	1HzHp Epoch-EBAR	17.49	1.34	-3.83	20	-5.91	1.54	0.003*
	1HzHp ICA-EBAR	13.9	0.99	1HzHp Epoch-EBAR	17.49	1.34	-3.55	20	-3.59	1.01	0.01*
	0.5HzHp No-EBAR	13.08	1.13	0.5HzHp ICA-EBAR	17.06	1.22	-3.33	20	-3.98	1.19	0.01*
	0.5HzHp No-EBAR	13.08	1.13	0.5HzHp Epoch-EBAR	19.12	1.75	-3.31	20	-6.03	1.82	0.01*
	0.5HzHp ICA-EBAR	17.06	1.22	0.5HzHp Epoch-EBAR	19.12	1.75	-1.62	20	-2.05	1.27	0.36
	0.1HzHp No-EBAR	15.76	1.54	0.1HzHp ICA-EBAR	20.33	1.57	-3.49	20	-4.57	1.31	0.01
	0.1HzHp No-EBAR	15.76	1.54	0.1HzHp Epoch-EBAR	19.77	2.49	-2.08	20	-4.02	1.93	0.15
	0.1HzHp ICA-EBAR	20.33	1.57	0.1HzHp Epoch-EBAR	19.77	2.49	0.27	20	0.56	2.06	1

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, p<0.05.

The second significant interaction is between lowpass/notch x EBAR. The main findings of post hoc comparisons between lowpass/notch x EBAR are summarised in table 4.10. It is found that, within EBAR and between lowpass/notch, using a 30HzLp

results in significantly lower peak amplitude than both 50Hz and NoLp/NF across all EBAR methods, as found in the main effect of EBAR. Moreover, comparisons within lowpass/notch and between EBAR find utilising no EBAR results in significantly lower peak amplitudes than ICA-EBAR and Epoch-EBAR across all lowpass/notch methods, matching the findings of the main effect of lowpass/notch method.

Table 4.10: Post hoc comparisons within the significant interaction of EBAR and lowpass/notch filter within the P300 peak amplitude ANOVA analysis. n=21.

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
EBAR Lowpass/Notch	No-EBAR NoLP/NF	14.41	1.11	No-EBAR 50HzNF	13.67	1.1	1.95	20	0.73	0.37	0.2
	No-EBAR NoLP/NF	14.41	1.11	No-EBAR 30HzLp	12.35	1.07	5.92	20	2.06	0.35	**<0.001***
	No-EBAR 50HzNF	13.67	1.1	No-EBAR 30HzLp	12.35	1.07	9.77	20	1.33	0.14	**<0.001***
	ICA-EBAR NoLP/NF	17.97	1.19	ICA-EBAR 50HzNF	17.31	1.13	2.4	20	0.67	0.28	0.08
	ICA-EBAR NoLP/NF	17.97	1.19	ICA-EBAR 30HzLp	16.02	1.07	7.32	20	1.96	0.27	**<0.001***
	ICA-EBAR 50HzNF	17.31	1.13	ICA-EBAR 30HzLp	16.02	1.07	9.13	20	1.29	0.14	**<0.001***
	Epoch-EBAR NoLP/NF	20.25	2.02	Epoch-EBAR 50HzNF	19.24	1.74	1.76	20	1.01	0.58	0.28
	Epoch-EBAR NoLP/NF	20.25	2.02	Epoch-EBAR 30HzLp	16.9	1.7	5.52	20	3.35	0.61	**<0.001***
	Epoch-EBAR 50HzNF	19.24	1.74	Epoch-EBAR 30HzLp	16.9	1.7	10.55	20	2.34	0.22	**<0.001***
Lowpass/Notch EBAR	NoLP/NF No-EBAR	14.41	1.11	NoLP/NF ICA-EBAR	17.97	1.19	-3.63	20	-3.57	0.98	**0.01***
	NoLP/NF No-EBAR	14.41	1.11	NoLP/NF Epoch-EBAR	20.25	2.02	-3.16	20	-5.84	1.85	**0.02***
	NoLP/NF ICA-EBAR	17.97	1.19	NoLP/NF Epoch-EBAR	20.25	2.02	-1.56	20	-2.27	1.46	0.41
	50HzNF No-EBAR	13.67	1.1	50HzNF ICA-EBAR	17.31	1.13	-3.42	20	-3.63	1.06	0.01
	50HzNF No-EBAR	13.67	1.1	50HzNF Epoch-EBAR	19.24	1.74	-3.23	20	-5.56	1.72	**0.01***
	50HzNF ICA-EBAR	17.31	1.13	50HzNF Epoch-EBAR	19.24	1.74	-1.47	20	-1.93	1.31	0.47
	30HzLp No-EBAR	12.35	1.07	30HzLp ICA-EBAR	16.02	1.07	-3.63	20	-3.67	1.01	**0.01***
	30HzLp No-EBAR	12.35	1.07	30HzLp Epoch-EBAR	16.9	1.7	-2.86	20	-4.55	1.59	**0.03***
	30HzLp ICA-EBAR	16.02	1.07	30HzLp Epoch-EBAR	16.9	1.7	-0.69	20	-0.88	1.27	1

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, $p<0.05$.

4.3.3.4) P300 Mean Amplitude

For the P300 mean amplitude analysis within the Cz electrode, Mauchly's test of Sphericity is violated for the main effects of highpass ($\chi^2(2)=31.08$, $p\leq0.001$, $\varepsilon=0.554$) and lowpass/notch ($\chi^2(2)=22.994$, $p\leq0.001$, $\varepsilon=0.588$). Moreover, the assumption of sphericity is violated for the interactions between highpass x lowpass/notch ($\chi^2(9)=40.429$, $p\leq0.001$, $\varepsilon=0.463$, highpass x EBAR ($\chi^2(9)=32.747$, $p\leq0.001$, $\varepsilon=0.535$), lowpass/notch x EBAR ($\chi^2(9)=47.379$, $p\leq0.001$, $\varepsilon=0.582$), and highpass x lowpass/notch x EBAR ($\chi^2(35)=476.401$, $p\leq0.001$, $\varepsilon=0.346$). All sphericity violations were adjusted with the Greenhouse-Geisser correction. The significant main effects and interactions are summarised in table 4.11.

Table 4.11: *Summary of the main effects and interactions of the P300 mean amplitude analysis. n=21.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Highpass	9.14	1.11	22.16	**0.01***	0.31
Lowpass/Notch	24.04	1.18	23.5	**<0.001***	0.55
EBAR	6.19	2	40	**0.01***	0.24
Highpass x Lowpass/Notch	4.38	1.85	37.05	**0.02***	0.18
Highpass x EBAR	5.99	2.14	42.76	**0.004***	0.23
Lowpass/Notch x EBAR	6.6	2.33	46.56	**0.002***	0.25
Highpass x Lowpass/Notch x EBAR	2.12	2.77	55.38	0.11	0.1

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

Within the P300 mean amplitude ANOVA, the main effect of all 3 factors reached significance. Post hoc comparisons within highpass found that using a 1HzHp filter resulted in a significantly smaller mean amplitude than 0.5HzHp and 0.1HzHp. Within lowpass/notch filters, using a 50HzNF resulted in a significantly larger mean amplitude than both NoLp/NF and 30HzLp. Within EBAR, the only significant

difference was between ICA-EBAR (M=9.9 ± 0.8) and No-EBAR, with ICA-EBAR resulting in a larger mean amplitude.

The interaction between highpass x lowpass/notch was significant, and the post hoc comparisons are summarised in table 4.12. Post hoc comparisons of the interaction between highpass x lowpass/notch find that the significant differences do not deviate from the main effects of both factors. Within lowpass/notch and between highpass, using a 1HzHp filter significantly reduces mean amplitude compared to both 0.5HzHp and 0.1HzHp filters across all lowpass/notch filtering methods. Within highpass and between lowpass/notch, using the 50Hz notch filter results in significantly larger mean amplitudes than NoLp/NF and 30HzLp across each highpass filtering method.

Table 4.12: *Post hoc comparisons within the significant interaction of highpass filter and lowpass/notch filter within the P300 mean amplitude ANOVA analysis. n=21.*

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
Lowpass/Notch Highpass	NoLP/NF 1HzHp	6.19	0.57	NoLP/NF 0.5HzHp	8.1	0.82	-4.7	20	-1.9	0.41	**<0.001***
	NoLP/NF 1HzHp	6.19	0.57	NoLP/NF 0.1HzHp	10.02	1.33	-3.17	20	-3.83	1.21	**0.01***
	NoLP/NF 0.5HzHp	8.1	0.82	NoLP/NF 0.1HzHp	10.02	1.33	-2.06	20	-1.93	0.94	0.16
	50HzNF 1HzHp	6.39	0.59	50HzNF 0.5HzHp	8.39	0.83	-4.97	20	-2	0.4	**<0.001***
	50HzNF 1HzHp	6.39	0.59	50HzNF 0.1HzHp	10.38	1.3	-3.43	20	-4	1.17	**0.01***
	50HzNF 0.5HzHp	8.39	0.83	50HzNF 0.1HzHp	10.38	1.3	-2.23	20	-2	0.9	0.11
	30HzLp 1HzHp	6.24	0.58	30HzLp 0.5HzHp	8.13	0.82	-4.63	20	-1.89	0.41	**<0.001***
	30HzLp 1HzHp	6.24	0.58	30HzLp 0.1HzHp	10.02	1.34	-3.1	20	-3.78	1.22	**0.02***
	30HzLp 0.5HzHp	8.13	0.82	30HzLp 0.1HzHp	10.02	1.34	-2	20	-1.9	0.95	0.18
Highpass Lowpass/Notch	1HzHp NoLP/NF	6.19	0.57	1HzHp 50HzNF	6.39	0.59	-6.3	20	-0.19	0.03	**<0.001***
	1HzHp NoLP/NF	6.19	0.57	1HzHp 30HzLp	6.24	0.58	-2.04	20	-0.05	0.02	0.16
	1HzHp 50HzNF	6.39	0.59	1HzHp 30HzLp	6.24	0.58	4.07	20	0.14	0.04	**0.002***
	0.5HzHp NoLP/NF	8.1	0.82	0.5HzHp 50HzNF	8.39	0.83	-5.38	20	-0.29	0.05	**<0.001***
	0.5HzHp NoLP/NF	8.1	0.82	0.5HzHp 30HzLp	8.13	0.82	-1.29	20	-0.03	0.02	0.64
	0.5HzHp 50HzNF	8.39	0.83	0.5HzHp 30HzLp	8.13	0.82	4.41	20	0.26	0.06	**0.001***
	0.1HzHp NoLP/NF	10.02	1.33	0.1HzHp 50HzNF	10.38	1.3	-4.13	20	-0.36	0.09	**0.002***
	0.1HzHp NoLP/NF	10.02	1.33	0.1HzHp 30HzLp	10.02	1.34	-0.09	20	0	0.03	1
	0.1HzHp 50HzNF	10.38	1.3	0.1HzHp 30HzLp	10.02	1.34	3.83	20	0.36	0.09	**0.003***

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, $p<0.05$.

The interaction between highpass x EBAR reached significance for the P300 mean amplitude (table 4.13). Post hoc comparisons within EBAR and between highpass found only comparisons within ICA-EBAR methods significantly differed, with ICA-EBAR 0.1HzHp resulting in larger mean amplitudes than both ICA-EBAR 1HzHp and ICA-EBAR 0.5HzHp. Moreover, the mean amplitude when using ICA-EBAR 0.5HzHp was significantly larger than when using ICA-EBAR 1HzHp. Within highpass and between EBAR, it is found that utilising ICA-EBAR resulted in significantly higher mean amplitudes than No-EBAR across all highpass filters, matching the main effect of EBAR. Moreover, an additional comparison reaches significance between EBAR methods within the 1HzHp highpass filters only, with 1HzHp Epoch-EBAR resulting in a larger mean amplitude than 1HzHp No-EBAR.

Table 4.13: *Post hoc comparisons within the significant interaction of EBAR and highpass filter within the P300 mean amplitude ANOVA analysis. n=21.*

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
EBAR Highpass	No-EBAR 1HzHp	4.47	0.82	No-EBAR 0.5HzHp	5.75	1.05	-2.48	20	-1.29	0.52	0.07
	No-EBAR 1HzHp	4.47	0.82	No-EBAR 0.1HzHp	8.15	1.51	-2.55	20	-3.68	1.45	0.06
	No-EBAR 0.5HzHp	5.75	1.05	No-EBAR 0.1HzHp	8.15	1.51	-2.29	20	-2.39	1.05	0.1
	ICA-EBAR 1HzHp	6.88	0.58	ICA-EBAR 0.5HzHp	9.97	0.95	-6.82	20	-3.09	0.45	**<0.001***
	ICA-EBAR 1HzHp	6.88	0.58	ICA-EBAR 0.1HzHp	12.88	1.24	-5.91	20	-6.01	1.02	**<0.001***
	ICA-EBAR 0.5HzHp	9.97	0.95	ICA-EBAR 0.1HzHp	12.88	1.24	-2.75	20	-2.92	1.06	**0.04***
	Epoch-EBAR 1HzHp	7.48	0.86	Epoch-EBAR 0.5HzHp	8.89	1.19	-2.4	20	-1.41	0.59	0.08
	Epoch-EBAR 1HzHp	7.48	0.86	Epoch-EBAR 0.1HzHp	9.39	1.9	-1.24	20	-1.91	1.54	0.69
	Epoch-EBAR 0.5HzHp	8.89	1.19	Epoch-EBAR 0.1HzHp	9.39	1.9	-0.48	20	-0.51	1.06	1
Highpass EBAR	1HzHp No-EBAR	4.47	0.82	1HzHp ICA-EBAR	6.88	0.58	-2.84	20	-2.41	0.85	**0.03***
	1HzHp No-EBAR	4.47	0.82	1HzHp Epoch-EBAR	7.48	0.86	-3.1	20	-3.01	0.97	**0.02***
	1HzHp ICA-EBAR	6.88	0.58	1HzHp Epoch-EBAR	7.48	0.86	-0.79	20	-0.6	0.76	1
	0.5HzHp No-EBAR	5.75	1.05	0.5HzHp ICA-EBAR	9.97	0.95	-3.67	20	-4.21	1.15	**0.01***
	0.5HzHp No-EBAR	5.75	1.05	0.5HzHp Epoch-EBAR	8.89	1.19	-2.38	20	-3.13	1.32	0.08
	0.5HzHp ICA-EBAR	9.97	0.95	0.5HzHp Epoch-EBAR	8.89	1.19	0.99	20	1.08	1.09	1
	0.1HzHp No-EBAR	8.15	1.51	0.1HzHp ICA-EBAR	12.88	1.24	-3.7	20	-4.74	1.28	**0.004***
	0.1HzHp No-EBAR	8.15	1.51	0.1HzHp Epoch-EBAR	9.39	1.9	-0.9	20	-1.24	1.38	1
	0.1HzHp ICA-EBAR	12.88	1.24	0.1HzHp Epoch-EBAR	9.39	1.9	2.02	20	3.49	1.73	0.17

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, $p<0.05$.

The third interaction to reach significance within the P300 mean amplitude analysis is lowpass/notch x EBAR, which is summarised in table 4.14. Post hoc comparisons within EBAR and between lowpass/notch find that the mean amplitude when using

No-EBAR 30HzLp is significantly larger than No-EBAR NoLp/NF, ICA-EBAR 50HzNF is significantly larger than both ICA-EBAR NoLp/NF & ICA-EBAR 30HzLp, and Epoch-EBAR 50HzNF is significantly larger than both Epoch-EBAR NoLp/NF & Epoch-EBAR 30HzLp. Post hoc comparisons within lowpass/notch and between EBAR find that using ICA-EBAR results in significantly larger mean amplitudes than No-EBAR across all lowpass/notch filtering methods, mirroring significant comparisons within the main effect of EBAR.

Table 4.14: *Post hoc comparisons within the significant interaction of EBAR and lowpass/notch filter within the P300 mean amplitude ANOVA analysis. n=21.*

	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	p
EBAR Lowpass/Notch	No-EBAR NoLP/NF	6.06	0.98	No-EBAR 50HzNF	6.22	1	-2.17	20	-0.16	0.08	0.13
	No-EBAR NoLP/NF	6.06	0.98	No-EBAR 30HzLp	6.09	0.98	-2.62	20	-0.03	0.01	**0.049***
	No-EBAR 50HzNF	6.22	1	No-EBAR 30HzLp	6.09	0.98	1.69	20	0.14	0.08	0.32
	ICA-EBAR NoLP/NF	9.83	0.81	ICA-EBAR 50HzNF	10.05	0.83	-3.35	20	-0.22	0.07	**0.01***
	ICA-EBAR NoLP/NF	9.83	0.81	ICA-EBAR 30HzLp	9.84	0.8	-0.28	20	-0.01	0.05	1
	ICA-EBAR 50HzNF	10.05	0.83	ICA-EBAR 30HzLp	9.84	0.8	3.07	20	0.21	0.07	**0.02***
	Epoch-EBAR NoLP/NF	8.42	1.22	Epoch-EBAR 50HzNF	8.88	1.23	-7.58	20	-0.46	0.06	**<0.001***
	Epoch-EBAR NoLP/NF	8.42	1.22	Epoch-EBAR 30HzLp	8.46	1.23	-1.82	20	-0.04	0.02	0.25
	Epoch-EBAR 50HzNF	8.88	1.23	Epoch-EBAR 30HzLp	8.46	1.23	6.1	20	0.42	0.07	**<0.001***
Lowpass/Notch EBAR	NoLP/NF No-EBAR	6.06	0.98	NoLP/NF ICA-EBAR	9.83	0.81	-3.86	20	-3.77	0.98	**0.003***
	NoLP/NF No-EBAR	6.06	0.98	NoLP/NF Epoch-EBAR	8.42	1.22	-2.03	20	-2.36	1.16	0.17
	NoLP/NF ICA-EBAR	9.83	0.81	NoLP/NF Epoch-EBAR	8.42	1.22	1.29	20	1.41	1.1	0.64
	50HzNF No-EBAR	6.22	1	50HzNF ICA-EBAR	10.05	0.83	-3.75	20	-3.83	1.02	**0.004***
	50HzNF No-EBAR	6.22	1	50HzNF Epoch-EBAR	8.88	1.23	-2.21	20	-2.65	1.2	0.12
	50HzNF ICA-EBAR	10.05	0.83	50HzNF Epoch-EBAR	8.88	1.23	1.05	20	1.17	1.11	0.91
	30HzLp No-EBAR	6.09	0.98	30HzLp ICA-EBAR	9.84	0.8	-3.85	20	-3.76	0.98	**0.003***
	30HzLp No-EBAR	6.09	0.98	30HzLp Epoch-EBAR	8.46	1.23	-2.04	20	-2.37	1.17	0.17
	30HzLp ICA-EBAR	9.84	0.8	30HzLp Epoch-EBAR	8.46	1.23	1.27	20	1.39	1.09	0.66

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, p<0.05.

4.3.3.5) Summary of the Statistical Analysis

A summary of the statistical results is shown in Table 4.15, which shows that the effects of lowpass/notch filtering method and EBAR, along with the interactions

between highpass x EBAR and lowpass/notch x EBAR are significant for each component and analysis examined. The effect of highpass and interaction between highpass x EBAR were only significant for the P300 window, with the interaction only being significant for the mean amplitude analysis.

Table 4.15: *Summary of the results of the statistical analysis for the comparison of preprocessing steps. The ANOVA results for the N170 peak amplitude, N170 mean amplitude, P300 peak amplitude and P300 mean amplitude for each main effect and interaction. Main effects and interactions which reach significance p<0.001 are marked with **, main effects and interactions which reach significance p<0.05 are marked with *, and non-significant results are marked with 'ns'.*

	N170 Peak Amplitude	N170 Mean Amplitude	P300 Peak Amplitude	P300 Mean Amplitude
Highpass	ns	ns	*	*
Lowpass/Notch	**	**	**	**
EBAR	*	*	*	*
Highpass x Lowpass/Notch	ns	ns	ns	*
Highpass x EBAR	*	**	*	*
Lowpass/Notch x EBAR	*	*	*	*
Highpass x Lowpass/Notch x EBAR	ns	ns	ns	ns

4.4) Discussion

Three artifacts linked to the use of HMD-VR during EEG recordings were found within the HS-HMD-VR presented arithmetic task reported in Chapter 3: electrical line noise; eye-related artifacts including eyeblinks and eye movements; and sweat-related slow drift. Whilst the artifacts were reduced or removed using standard EEG preprocessing guidance in the arithmetic experiment, no existing comparison between HMD-VR or HS-HMD-VR preprocessing pipelines could be found. As HS-HMD-VR may exacerbate the recorded artifacts, the improper selection of preprocessing parameters may not sufficiently reduce EEG artifacts, or might result in over-attenuation of the data. Three preprocessing steps related to the removal of the identified EEG artifacts, highpass filtering, lowpass/notch filtering, and eye movement artifact removal, were examined to identify appropriate preprocessing

pipelines parameters for HS-HMD-VR/EEG data. To the researcher's knowledge, this is the first formal comparison of these preprocessing methods within the context of HMD-VR/EEG research, and offers an important contribution into understanding how EEG artifacts can be removed from the waveform.

The purpose of analysis in this Chapter was not to simply identify which combination of preprocessing steps produces the largest ERP peak and mean amplitudes, however the differences in amplitude between the pipelines does provide insight into how each processing step changes the data. The mean and peak amplitudes of two ERP components identified in Chapter 3, the N170 and P300, were analysed over separate 3x3x3 ANOVA for easy-visual epochs in the centro-medial electrode Cz.

The effects of lowpass/notch filter and EBAR reached significance for the N170 peak and mean amplitude ANOVAs, and all three main effects reached significance for the P300 peak and mean amplitude ANOVAs. Within highpass filters, using a 0.1Hz highpass filter resulted in the largest P300 components. Within lowpass/notch filters, using no lowpass resulted in the largest N170 mean and peak amplitudes, but the 50Hz notch filter resulted in the largest P300 mean amplitude. Using no lowpass filter or a 50Hz notch filter did not significantly differ the P300 peak amplitude, but both resulted in a larger peak amplitude than when using a 30Hz lowpass. Within EBAR methods, using EBAR-ICA resulted in the smallest N170 peak and mean amplitudes, but also the smallest P300 peak amplitude. Only EBAR-ICA resulted in a larger P300 mean amplitude over No-EBAR.

The differences in which preprocessing steps result in the largest mean or peak amplitude between N170 and P300 ERP components demonstrates there is not a one-size fits all solution for optimising ERP responses in HS-HMD-VR research. Moreover, the interactions between the processing steps can result in the order of amplitude sizes within comparisons of a one processing step changing, depending on what other processing steps are used in the same pipeline. For example, the N170 peak amplitude comparison between 1Hz and 0.5Hz highpass filters within

EBAR methods finds that, when using ICA-EBAR, 1Hz results in a significantly larger peak. However, when No-EBAR is used, using a 0.5Hz results in the larger N170 peak amplitude. Moreover, it is known that earlier steps can affect the success or effectiveness of later steps, for instance 1Hz filters supporting later ICA compared to lower frequency value highpass filters [321], though with the trade-off of having reduced ERP peaks [310]. Therefore, examination of each decision individually is required to identify effectiveness of the artifact removal method against the impact on the ERP waveform, and thus to identify what compromises were suitable for HS-HMD-VR/EEG data.

4.4.1) Eye-Based Artifact Removal Methods

In the Chapter 3 arithmetic experiment, eyeblinks and eye movement artifacts were found to be prominent in many of the datasets collected. When the epochs are averaged together, the eye-based artifacts from the data were found to be represented as a negative inflection within the ERP waveforms that encompassed the N170 and P300 time windows. The negative inflection was followed by a large positive inflection at ~500ms which persisted past 1000ms, and would encompass later components such as the SWC. To ensure the integrity of any analysis conducted on these components, the artifact must be removed.

In Chapter 3, ICA was used to remove eye-based artifacts, but other HMD-VR experiments have rejected eyeblink contaminated epochs to successfully remove the artifact [205]. In the current study, it is found removing eye movement artifacts through epoch rejection is unsuitable for the current HS-HMD-VR and task configuration. Of the 240 epochs across all four conditions included in the Chapter 3 arithmetic experiment for each dataset, only an average of 38% of epochs were accepted when eye motion artifacts were rejected. Moreover, six datasets contained conditions with less than six epochs remaining when epoch rejection was used to remove eye movement artifacts. As these datasets failed to meet the minimum threshold required for inclusion [282,284], an additional 6% of the total data would be removed from the final analysis. When the total number of rejected epochs is

compared to the 84% accepted when using ICA analysis, using epoch rejection becomes difficult to recommend for the data collected using HS-HMD-VR.

Unlike Epoch-EBAR, ICA-EBAR successfully removes eyeblinks without discarding large percentages of the data, resulting in a clean waveform that contains clear N170 and P300 components. The statistical analysis found that ICA-EBAR resulted in the smallest N170 peaks, which likely results from the full removal of eye-movement artifacts [322,323] in the EBAR-ICA condition, opposed to the ICA reducing the amplitude of the ERP component [284]. It has previously been demonstrated that removing both eyeblink and movement artifacts using ICA resulted in a negative trend within the ERP waveforms being fully removed [323]. In Figure 4.14, most filter combinations have a late negative inflection present in the Epoch-EBAR but not in the ICA-EBAR variants within the Pz electrodes. Therefore, it is possible that subtle eye movements which were not or could not be identified during visual inspection of the epochs were removed by the ICA automatic classification system, rendering ICA as the superior method for their removal. However, as the Epoch-EBAR variants were more positive towards the anterior of the brain, removing the negative trend in the waveform may be dependent on the electrode examined. Regardless of the reason, the comparisons conducted highlight the need for additional consideration when selecting what artifact rejection technique to use within HMD-VR environments, particularly if a task encouraging or requiring visual exploration of the VE is used which may provoke eye artifacts.

4.4.2) Highpass Filters

Comparisons between highpass filter frequency values were conducted to compare the removal of the slow wave artifacts present in the Chapter 3 arithmetic task datasets. Slow wave artifacts in EEG recordings can be caused by sweat-related changes in skin impedance, artificially increasing or decreasing the amplitude over time [290]. Sweat-related drift is a large concern during HMD-VR usage, as it has been reported perspiration can increase for some participants during HMD-VR usage [289]. In the Chapter 3 arithmetic study, ~30% of participants self-reported increased sweating as part of the SSQ during HS-HMD-VR usage. However, evidence of slow

drift artifacts is also found in participants who did not report increased perspiration (Figure 4.6), suggesting an alternative cause of drift, such as the conductive electrode gel drying [294].

For the P300 ERP component, it was found that using a 1Hz highpass filter results in a significantly smaller peak and mean amplitude than 0.5Hz and 0.1Hz highpass filters, which persists when restricting datasets to ICA-EBAR data only. When restricting the data to ICA-EBAR datasets, it is also found that utilising higher frequency value highpass filters reduces the amplitude of P300 peaks, as expected from previous comparisons between highpass filters [45,290,306,310]. Moreover, the relative increase of N170 amplitude when utilising higher frequency highpass filters matches those found by Kappenman & Luck [290] within a warm and humid environment, similar to what may be experienced during HMD-VR usage. The removal of the increased negativity of the N170 indicates that positive slow-wave contamination has been removed when using the 0.5Hz and 1Hz highpass filters.

It is found in this study that use of higher frequency value highpass filtering results in an overall reduced positivity within the averaged ERP waveform, resulting in a smaller P300 but larger N170 response. Whilst it has been argued by Luck [45] and Tanner et al. [310] that the use of filters above 0.1Hz can result in distorted waveforms, evidence collected within HS-HMD-VR data suggests that a 0.1Hz highpass filter is not sufficient to fully remove slow drift artifacts recorded during combined EEG and HS-HMD-VR usage. Conversely, applying a higher frequency value highpass filter than necessary increases the amplitude of earlier negative components such as the N170 at the cost of reducing the amplitude of the later positive P300 component. It is therefore argued here that using a 0.5Hz highpass filter offers the optimal trade-off between minimising the slow wave artifacts and distorting the ERP. The use of the 0.5Hz highpass filter offers the additional benefit of being suitable for multiple ERP components, for example the N170 and P300, whilst not over-modulating either during analysis. Moreover, it has been previously reported that using 0.5Hz highpass filter increases the statistical power of the data compared to 0.1Hz highpass filtered data [290,298], providing an additional benefit.

4.4.3) Lowpass and Notch Filtering

The third factor examined is the lowpass/notch filter, which in the current preprocessing pipeline is used to remove high-frequency electrical line noise from the EEG recording resulting from the Vive Pro HS-HMD-VR device [56,287]. Within HMD-VR research, line noise has previously been removed using lowpass filters set below the line noise frequency [16], notch filters targeting 50Hz or 60Hz line noise frequencies [64], or has not been removed from the EEG data [205].

The statistical analysis conducted revealed that using the 30HzLp resulted in significantly smaller N170 and P300 Cz peak amplitudes than using a 50HzNF or no lowpass filter. Moreover, using no lowpass or notch filter resulted in a significantly larger N170 peak amplitude compared to the 50Hz notch filter, but not between P300 peak amplitudes. Mean amplitude comparisons found that using no lowpass resulted in a larger N170 mean amplitude compared to 50HzNf and 30Hz, whereas the 50HzNF resulted in a significantly larger mean P300 compared to no lowpass and 30Hz.

In the context of ERPs captured within HS-HMD-VR, the close-proximity of the recording EEG cap to the HMD renders it difficult to recommend taking no steps to remove electrical artifacts introduced by the device. However, unless a researcher wants to remove the harmonics of the line noise, it is also unnecessary to utilise a stricter lowpass filter. It is therefore argued that utilising a 50Hz notch filter targeting the frequency of the local line noise is a suitable compromise for the negatable reduction in peak amplitude.

4.5) Conclusions

The purpose of this chapter was not to imply the existence of a singular ideal EEG preprocessing pipeline universally applied to every dataset using any form of HMD-VR. Instead, the study aimed to prevent errors which could lead to misrepresentation of the data collected in HS-HMD-VR experiments. A balance between strictness and

leniency must be found in all EEG data preprocessing decisions, for example applying too strong filtering will over-attenuate the data, but using too weak artifact removal may result in false positive or negative differences. To this end, this comparison concludes that utilising a 0.5Hz highpass filter, a notch filter targeting local line noise, and ICA-based rejection of eyeblinks and eye motion artifacts is suitable for ERP data collected in a HS-HMD-VR experiment. The combination of preprocessing methodology identified allows the targeting of multiple ERP components of different polarities, whilst minimising slow wave drift, electrical line noise, and eye-movement related artifacts found in HS-HMD-VR EEG data. To date, and as far as the researcher is aware, this is the first comparison of EEG preprocessing methodologies for artifact removal in an EEG data collected in a HS-HMD-VR experiment. The recommendations for removing noise proposed in this chapter are formed from data acquired when using a simple HS-HMD-VR VE that required no movement during the ERP recording periods. The noise removed thus resulted from the use of HMD-VR rather than the task or VE used. Therefore, it is argued that the recommendations are not specific to arithmetic tasks, and should be applicable to experiments presented in HS-HMD-VR (at least those using the Vive Pro) which target the N170 or P300 component. Additional consideration must be given to other potential sources of noise depending on the task, such as participant movement during recording or originating from different HMD-VR devices. However, it is anticipated that the same or similar artifact removal methods to those described in this chapter should also be sufficient for reducing these potential sources of noise.

Chapter 5) A Comparison Learning and Recall in a Spatial Navigation Working Memory Maze Task Presented in Head-Mounted Display and Desktop Virtual Reality Environments

5.1) Introduction

The experiment conducted in Chapter 3 demonstrated that HS-HMD-VR and EEG can be combined to successfully acquire ERP responses during working memory arithmetic tasks. The analysis of ERP responses in working memory studies typically focuses on ERP components occurring within the first 1000ms post-stimuli presentation, such as the ~140-210ms N170, the ~200-550ms P3/P300 and the ~450-750ms post-P300 SWC components identified in Chapter 3. Although ERP responses in working memory studies have been examined up to 4000ms post-stimulus onset [324], there are many working memory-related cognitive processes, including sustained attention [325], mental manipulation/rotation [326] and spatial navigation [327], which can persist past four seconds.

5.1.1) Electroencephalography Frequency Analysis using Power Spectral Density

An alternative methodology for analysing EEG data is frequency analysis, which examines changes within bands of neuronal oscillations associated with cognitive processes [328–330]. Neurons in brain regions associated with a given cognitive process can synchronise or desynchronise during either activation or inhibition (depending on the cognition, brain region, and oscillation frequency), resulting in measurable changes in the amplitudes of different frequency bands [331,332]. The changes in amplitude can then be compared to a baseline or another condition to gain a deeper insight into neural processing. Frequency analysis methods can be time-locked for examining changes in frequency over time, such as event-related spectral perturbation ERSP [333] and ERD/ERS [334].

Frequency analysis can also be examined over non-time locked events using a power spectral density (PSD) analysis, which in EEG analyses typically utilises the 'Welch method' [49,330,335]. The Welch method uses a series of discrete Fourier transforms, which convert data from the time domain to the frequency domain [331]. These discrete Fourier transforms are typically conducted over short overlapping time windows for the whole time series data to smooth the data, before being averaged together to produce the PSD [49,330]. From this PSD, the absolute power or relative power between trial and a baseline for selected frequency bands can be extracted and statistically analysed [336,337]. In addition to analysing the power of individual frequency bands, ratios between absolute power can also be utilised. For example, the theta/alpha ratio (TAR), also called the 'cognitive load index' or 'task load index', is a ratio between the frontal theta and parietal alpha activity, typically in the fronto-medial and parietal-medial electrodes, and is used to calculate differences in WML between conditions [129,338].

5.1.2) Power Spectral Density Frequency Analysis in Working Memory High-Specification Head-Mounted Display Virtual Reality Research

Frequency analysis of EEG data has previously been utilised in HMD-VR experiments. For example, collecting ERD data using the non-modern virtual Research Eight HMD [339]; ERSP data using smartphone-HMD-VR [340]; PSD using the development-grade Oculus DK1 [341], or combinations of the PSD, ERD and ERSP using the consumer-grade Oculus CV1 [342,343]. Several examples of frequency analysis comparing WML between HMD-VR and an alternative display method were identified in the systematic review conducted in Chapter 2. For example, Juliano et al. [175] examined relative beta activity, Li et al. [177] used a beta/theta ratio, and Škola & Liarokapis [179] utilised a classifier to distinguish WML using mu & beta data in modern HMD-VR experiments.

To date there has been little working memory or adjacent research using the Vive Pro HS-HMD-VR which analysed the data using absolute or relative frequency power analysis. Of the few examples identified, the primary focus has been on measuring differences in WML between conditions presented within the HS-HMD-VR

VEs. Most studies identified compared changes in frontal theta and/or parietal alpha frequency bands [344–346], one study compared the relative power of, and a ratio between, frontal and occipital beta alpha [67], and one utilised multivariate pattern classification methods for theta, alpha and beta activity [347]. The focus on the theta, alpha and beta frequency bands is not surprising, as these are the most commonly reported frequencies in the WML literature [122], yet there are some important gaps in the HS-HMD-VR literature. Primarily, there is a distinct lack of comparative working memory studies utilising PSD-based power analysis between HS-HMD-VR and an alternative display in working memory studies.

5.1.3) Spatial Navigation and Working Memory

One area of working memory that has taken advantage of the immersive VEs offered by HMD-VR devices, and would also facilitate frequency analysis over longer trial durations to compare levels of WML, is spatial navigation. Spatial navigation is the cognitive process of establishing, maintaining and updating a route between two points in space, either through recall or cues in the environment. Spatial navigation relies on working memory processes for the maintenance and updating of route information, for example attention [348], decision making [349], and executive function [350]. Spatial navigation is further associated with visuospatial processes and visuospatial working memory (VSWM) for processing external cues and information about an individual's current location relative to a goal or target [351–353]. Spatial navigation is a vital process that humans undertake in daily life, such as navigating through a supermarket whilst shopping [354] or changing lanes whilst driving a car [355]. However, increasing working memory demands during navigation can negatively impact performance and later route recall [356].

Spatial processing can be divided between egocentric and allocentric processing, which are explored through variations of spatial navigation tasks. Egocentric spatial processing is where locations of external points in space are processed relative to the individual's current perspective [357,358]. Egocentric spatial navigation is commonly investigated using route learning tasks, which involve guiding participants through an environment with instruction to remember the navigation steps, before

having them retravel the route unassisted [359]. Allocentric spatial processing is independent of the individual, and is where the representation of external points in space are processed relative to each other [358,360]. In research, allocentric tasks include learning navigation routes using maps [361,362], but can include task performance measures following egocentric presentation. For example, completing a cognitive map from memory following egocentric exploration [363,364]. Both egocentric and allocentric spatial navigation have been linked to visuospatial working memory, and those with a higher VSWM capacity have been reported to perform better in spatial navigation tasks [351,361,362,365].

Spatial navigation tasks will often include neurophysiological measures to examine neural correlates of spatial navigation or VSWM. Primarily, EEG is used to examine alpha, beta, and theta neural oscillations through power analysis or ERD/ERS during navigation [70,366–368]. The most commonly reported frequency band for spatial navigation, theta, is found to increase in the frontal region during more complex navigation and during recall of previously learnt routes [369,370]. In the context of spatial navigation, theta activity is believed to be associated with the processing and encoding of route navigation information [371], along with the working memory processes of learning and decision making [372]. Similarly, attention-related parietal alpha activity during navigation has been found, though this has been reported to both increase or decrease depending on the task [370,372].

5.1.3.1) Spatial Navigation and Head-Mounted Display Virtual Reality

There has been particular interest in the use of HMD-VR in spatial navigation research and application, as the displays offer high ecological validity, presence, and embodiment within VEs [373]. HMD-VR has been utilised in a range of spatial navigation paradigms, including spatial memory [374], spatial updating [375,376], the effect of spatial cognitive training on route learning [377], and the use of landmarks during route learning [378]. HS-HMD-VR Vive Pro devices have also seen use in spatial navigation research in combination with EEG [379,380].

The high ecological validity and benefits offered by HMD-VR in spatial navigation is demonstrated through the similarities with real-life navigation, and the improvement over DB-VR navigation. Murcia-López & Steed [146] found that participants in an HMD-VR VE emulated the navigation paths used in real-life navigation, and had better spatial recall over DB-VR. Ruddle et al. [381] compared DB-VR and non-modern DB-HMD-VR recall following route learning through a building. They found that participants not only navigated faster in the HMD-VR condition, but also visually explored the environment more. The difference in visual exploration can be partially attributed to the keyboard and mouse compared to the naturalistic head-turning used in the DB-VR and HMD-VR conditions respectively. The differences between inputs highlights the importance of balancing the input methods for accurate comparison.

Despite the advantages offered by HMD-VR, not every study has reported that HMD-VR enhances spatial navigation performance relative to other displays. Marraffino et al. [382] had participants complete a scavenger hunt in HMD-VR and DB-VR, and found no difference in the level of detail between the allocentric cognitive maps, which are the drawn recreations of the navigated route recalled from memory, produced between display conditions. However, they did report that participants with more video game experience performed better in the HMD-VR condition. A similar lack of difference between cognitive maps produced post-exploration was found in the comparison of DB-HMD-VR and real-life environment by Dong et al. [383].

Some papers have found that spatial navigation performance in HMD-VR can be reduced relative to DB-VR presentation. Plechatá et al. [354] tasked participants with exploring a virtual supermarket in either DB-VR or HMD-VR, and later navigating from memory to a set of target items. Older participants performed better in the DB-VR condition, and whilst younger participants did not differ between displays. It was also found that higher levels of fatigue were reported during HMD-VR usage, suggesting cybersickness may be a concern in HMD-VR based spatial navigation. Moreover, Srivastava et al. [166] compared cognitive maps following exploration of an urban VE presented using HMD-VR and DB-VR. Cognitive maps were found to

be more accurate in the DB-VR condition, whilst inducing a lower WML and cybersickness scores than HMD-VR.

5.1.4) Maze Tasks in Working Memory Research

The use of maze tasks are common in the spatial navigation and working memory literature, offering a controlled way of presenting egocentric learning and recall paradigm in the context of spatial navigation [384]. For example, a typical learning and recall maze paradigm will guide participants through a maze using visual cues at junctions, which the participant must attend to and hold in working memory. The visual cues are then removed on later repetitions, which the participant must navigate unguided from memory [70,370].

Maze tasks also facilitate the easy integration and testing of additional elements that may affect spatial navigation. A simple route learning task between a set start and finish point [369,385] can be integrated with landmarks at junctions [370] or on the walls of the maze [386] to provide additional visual cues. Moreover, mazes can be easily adapted to provide different levels of WML through increasing or decreasing the length and number of turns within a maze [369,387]. Alternatively, difficulty can be increased or decreased by adding useful or distracting elements to the environment, for example by changing the colours of the walls [70]. Secondary tasks have similarly been utilised in maze paradigms to increase experienced WML during navigation [388,389]. As with other forms of spatial navigation, Meneghetti et al. [388] demonstrated that higher VSWM capacity or ability improves task performance during maze navigation when under higher levels of load.

There are several different types of maze layouts used by researchers which are designed for different tasks, and are largely divided between radial arm mazes, traditional mazes, and T-junction mazes. Radial arm mazes [390] are simple central circular rooms with several corridors branching off, one of which will be correct (Figure 5.1a). Radial arm mazes are typically used in the study of learning and memory in rats [391,392], but have more recently been proposed to be used in

human research through VR [327,390]. Traditional mazes (such as those used by Erkan [393] and Sharma et al. [394]) are more akin to real-world mazes, containing start and end points and comprising of inconsistent pathway lengths and turns (Figure 5.1b). Traditional mazes place a heavier emphasis on exploration but can introduce a wealth of secondary information that may inform participants about the correct path. For example, a participant may remember a turn halfway down a corridor following the first left turn, as opposed to remembering just the sequence of turns.

(a) Radial maze

(b) Traditional maze

(c) T-junction maze

(d) 3-Choice T-junction maze

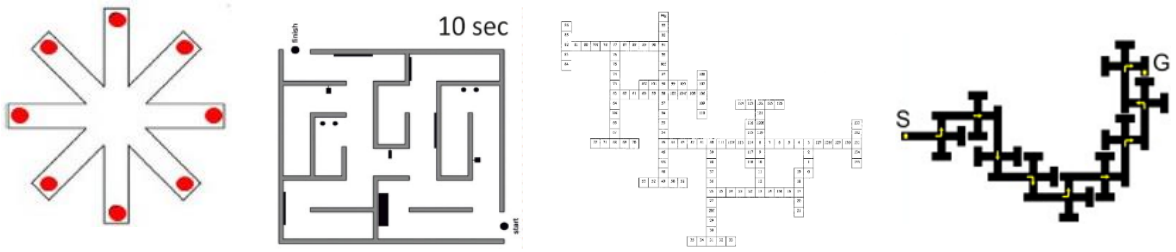


Figure 5.1: Examples of types of mazes used in previous spatial navigation paradigms. *Palombi et al. [390]* used a radial maze (a); *Erkan [393]* used a traditional maze (b); Bischof & Boulanger [70] used a T-junction maze (c); and *Kober et al. [370]* used a 3-choice variant of the T-junction maze (d).

T-junction mazes, such as those used by Bischof & Boulanger. [70], Kahana et al. [369] and Kober et al. [370], are uniformed decision-based paths which are designed to minimise secondary information. Participants navigate down a corridor and are met with a set number of potential pathways to follow, leading to the next turn decision or a dead end. T-junctions will typically consist of binary 2-decision junctions (Figure 5.1c), but can utilise 3 or more in certain paradigms (Figure 5.1d).

Despite the interest in HMD-VR in spatial navigation, the use of mazes and guided navigation tasks are currently underrepresented in the HMD-VR literature. Whilst maze-based environments have previously been used in HMD-VR [394–397], very few investigate spatial navigation and outcomes from route learning. Of those that could be identified, only two examined route learning in comparison with another

display method. Sousa Santos et al. [398] compared HMD-VR and DB-VR in spatial navigation tasks performed in a maze, where participants had to recall the location of objects. They found that more objects were recalled in the DB-VR condition, along with faster navigation speed and further distance travelled. More recently, Hsieh et al. (2018) compared the completion time and post-exposure cognitive maps of a maze-based learning and recall task presented using DB-HMD-VR with motion controllers and a DB-VR using a keyboard and mouse. The mazes used consisted of unguided 3-choice junction mazes, with landmarks either placed in the distance, placed at each junction, or placed by the participant at will during navigation. It was found that mazes were completed quicker in the DB-VR condition, but the level of detail of the cognitive maps produced was higher in the HMD-VR condition.

5.1.4.1) Electroencephalography in Maze Navigation

A range of EEG methodologies have been utilised in maze navigation, finding changes in activation are primarily associated with the alpha and theta bands of activity [366,370]. It is typically found that frontal theta activity increases during spatial navigation of mazes [70,385], and is more frequently found during complex mazes over simple, and during recall over learning [369]. Theta has also been found to increase during active navigation relative to passive guided navigation [372]. Parietal alpha activity has also been found to increase during active navigation of mazes [372], but decrease when more immersive displays are used [370].

5.1.5) Aims of the Present Study

HMD-VR is poised to be one of the most important methodological advancements for spatial navigation research. The ability to present immersive fully controlled VEs during route learning, navigational recall, or free exploration could allow for a deeper exploration of human VSWM than previously possible.

To date, no published spatial navigation tasks utilising a learning and recall task within a complete T-junction maze in HS-HMD-VR could be identified. When considering the conflicting findings of other spatial navigation tasks comparing

between DB-VR and HMD-VR, it is currently unknown if HMD-VR would benefit or impair spatial navigation and working memory processes relative to DB-VR during maze navigation. Moreover, no comparison utilising EEG between HS-HMD-VR and DB-VR in a maze task was found, leaving it uncertain how experienced WML and associated neural responses during navigation differ between displays. To address this gap in the literature, and to increase understanding of how the use of HMD-VR impacts experienced WML relative to DB-VR during spatial navigation, a learning and recall task presented using HS-HMD-VR and DB-VR will be conducted. This learning and recall spatial navigation task will compare between the display methods, maze difficulty level and navigation types.

It is apparent throughout the literature that theta and alpha activation are heavily linked to both WML and spatial navigation, with the ratio between these frequencies serving as a measure of experienced cognitive load. Therefore, EEG power analysis of continuous data for theta and alpha bands of activity, in addition to the TAR, will be utilised to measure WML experienced between display conditions, between navigation types, and between maze difficulty through differing the length and number of turns.

Spatial navigation using HS-HMD-VR, and how it compares to other display methods, is an important yet underexplored aspect of the working memory literature. To date, no comparison of different levels of WML between HS-HMD-VR and an alternative display method during a full maze navigation task has been conducted. Moreover, it is unclear what effect different objectives during navigation, such as route learning or recall, will have on working memory processes between different display methods. This study aims to use a spatial navigation task to compare HS-HMD-VR and DB-VR using behavioural task performance and PSD-based EEG measures of WML. Frontal-central theta-band EEG activity, parietal-central alpha-band activity, and a theta-alpha ratio will be acquired during the active navigation of learning, active navigation of recall, and passive guided navigation of a path through a maze VE presented using HS-HMD-VR and DB-VR. WML will be manipulated by presenting mazes over 3 levels of increasing difficulty (4, 8 and 12 turns).

Behavioural measures, including participant task completion time, number of wrong turns, subjective difficulty rating, and cybersickness symptoms will be acquired. The absolute power of alpha and theta EEG frequency bands and the TAR will be compared between displays, difficulty level, and between learning, recall, and guided navigation using a repeated measures design.

5.2) Methods

5.2.1) Participants

A total of 27 participants aged between 18 and 55 (Mean Age ± S.E.M = 23.26 ± 0.9 years old, range: 18-39 years old) with normal-to-corrected vision were recruited (9 female, M= 22.8 ± 1.0 years old, range: 18-27 years old; 18 male, M= 23.5 ± 1.3 years old, range: 18-39 years old) through opportunity sampling at the University of Hull campus. The number of participants recruited was within the standard range for EEG research (see Larson and Carbine [183]). As it was unknown what navigation strategies would be employed by participants, the upper age limit is below the typical decline in allocentric spatial navigation ability associated with 60-70 year olds [399,400]. Each participant provided informed consent, and completed a paper-based health questionnaire (Appendix 5) to exclude individuals with medical concerns or neurological conditions that may affect cognitive function. Moreover, participants with medical conditions which necessitate an electrical stimulator, which may introduce noise to an EEG recording, were also excluded. The exclusion criteria were predefined by the researcher. The study was conducted in conformity with the Declaration of Helsinki [251], and received local ethical approval from the Hull York Medical School Ethics Committee (Reference 1303).

Based on the Edinburgh Handedness Inventory responses, 3 participants (2 male, 1 female) were left-handed. Four participants had never used HMD-VR before, 9 reported <1hour of total usage, 7 reported 1-10 hours, 4 reported 10-24 hours, and 2 reported over 24 hours of usage. Five participants reported playing an average of 0 hours of video games per week, 6 reported <1hour per week, 5 reported between 1-

7 hours per week, 5 reported 7-14 hours per week, and 2 reported over 14 hours a week.

A total of 7 participants exited the study before completion due to cybersickness. The early-exit participants comprised of all 4 who reported no prior HMD-VR experience, 1 participant who reported less than 1 hour of experience, and 2 participants with 1-10 hours of HMD-VR experience. Moreover, of the early-exit participants, 3 reported playing an average of 0 hours of video games a week, 2 reported playing less than 1 hour per week, and 1 reported playing between 1-7 hours per week.

Three participants were excluded due to technical difficulties with the EEG device preventing successfully data recording. Two participants were excluded due to not successfully completing at least one maze run of every trial block. One EEG dataset was removed during analysis due to high levels of unremovable noise. A total of 14 datasets were included in the analysis.

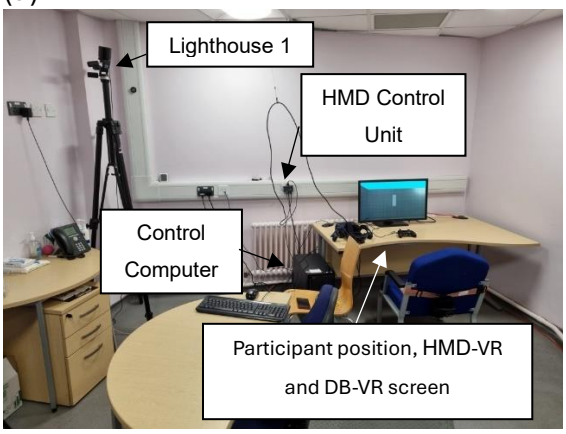
5.2.2) Materials and Apparatus

5.2.2.1) Hardware and Software

The experimental maze was constructed using the ‘Valve Hammer Editor’ software and presented using the ‘Garry’s Mod’ software. Two additional plugins for Garry’s Mod were used to support the experiment: The VR Mod to make the VE compatible with HMD-based 6DOF navigation; and the PyGmod module for compatibility with the Python-based code for the EEG triggers. The HMD-VR hardware was managed by SteamVR software, which was supplemented with the Driver4VR to allow the emulation of HMD-VR controller input using the Xbox 360 controller. Additional code was written in the Lua and Python programming languages to support the maze task. The Lua code controlled the maze, progression of the experiment, what the function of each square is, where the virtual avatars are teleported to, and the formatting and outputting of behavioural data. The Python code interfaced with the trigger wire to deliver EEG trigger marks to the to the Io:Bio recording device.

All experimental and VR software, along with display and input hardware, was managed by a desktop computer using an I7-9700K processor with a GTX2080 graphics card. The monitor used in the DB-VR condition was a 28 inch 3840 x 2160 (4K) pixel monitor with a 90Hz refresh rate. A HTC Vive Pro headset with a 1440 x 1600-pixel resolution, coupled with external tracking ‘lighthouses’ to track spatial position for 6 DOF exploration, was used in the HMD-VR condition. The HMD was calibrated for each participant by having the participant hold the HMD-VR in a comfortable position over the face, and then adjusting the head straps. Participants manually adjusted the FOV slider and IPD knob for visual clarity. The wire connecting the Vive Pro to the computer was trailed behind and to the side of the participant to prevent EEG headcap or trigger wires making contact. Audio was muted for both conditions, and the integrated Vive Pro headphones were kept off the ears. A standard wired Xbox 360 gamepad controller was used for input in both the HMD-VR and DB-VR display conditions. Rotating the point of view was possible in both condition using the right joystick on the gamepad controller, but participants could also move the point of view by turning their heads whilst wearing the HMD-VR. Participants were placed in a static seat to prevent participants physically turning at maze junctions. Mazes were navigated by participants using joystick-based continuous locomotion in both conditions. As participants were seated on a stationary chair, they had to rely on the controller to complete turns in the maze. All inputs on the controller were disabled to prevent participants unintentionally interrupting the experiment, except for the left and right joystick for lateral movement and camera rotation respectively. The ‘start’ button which opened the options menu could not be disabled, so participants were instructed to not touch this button. An additional keyboard and mouse were connected to the computer so the researcher could advance the experiment when required (Figure 5.2).

(a)



(b)

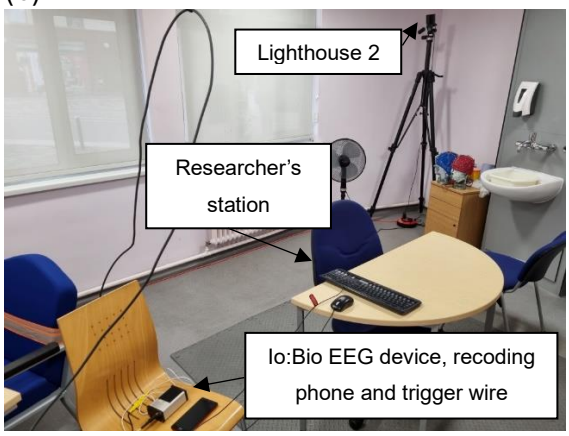


Figure 5.2: *Annotated images showing the configuration of the laboratory during the experimental procedure. (a) shows the participant position, HMD-VR configuration, DB-VR screen station. (b) shows the EEG system (Io:Bio) and the researcher's station. The researcher's station is rotated slightly to provide a better view of the participant station. The VE was calibrated to be centred around where the participant is seated. When not in use, the HMD-VR device was removed from the participant desk.*

5.2.2.2) Electroencephalography Recording

Spes Medica Sleepcaps with 19 tin electrodes (FP1, FP2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2) in the international 10-20 electrode configuration and a linked-ear reference was used to record the EEG signal. A headcap net was placed over the EEG cap when required to improve the connection between the electrodes and the scalp. The EEG headcap was connected to an Io:Bio physiological recording device [255], which in turn transmitted the data via Wi-Fi to an Asus Zenphone 6. EEG data was recorded using a 250Hz sampling rate. The Io:Bio device was also connected to the computer via a proprietary cable, which allowed it to receive EEG event triggers from the Python section of the maze function code.

5.2.2.3) Stimuli/Mazes

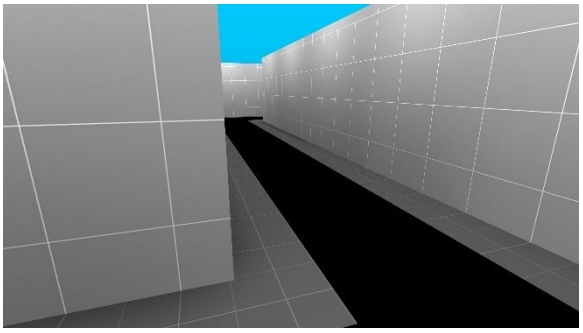
The grid-based T-junction mazes constructed for this experiment are based on the design principles as used by Bischof & Boulanger [70] and Kahana et al. [369]. The mazes were constructed using 'squares' placed into a square-grid template, with each grid space being the same size. Mazes were constructed from 6 types of square: 'Start Square', where the participant begins each trial; 'End Squares', which

if active mark the end of the trial and will teleport the participant's avatar to the start of the maze or to the preparation room; 'Correct Path' corridor squares with a single entrance and exit that are part of the correct route through the maze; 'Decision' corridor squares immediately before a junction; 'Junction' with an entrance and two exits, one left and one right; and 'Wrong Path' corridors or false junctions which lead to dead ends. Each section/turn follows the same format: participants walk down a 'correct path' corridor before reaching a 'decision' square. The square following the 'decision' is a 'junction', where they must make a binary left/right turn. Should the participant make the correct turn, they will then continue the maze and repeat the process until the end point is reached. Should the participant make a wrong turn, they will follow a 'wrong turn' corridor before being met with a dead-end at the false junction. Both turns following a false junction are immediately terminated with a wall after the space of one grid-square. The dead-end is not visible until participants reach the false junction, preventing visual cues indicating the correct decision being visible from the decision junction.

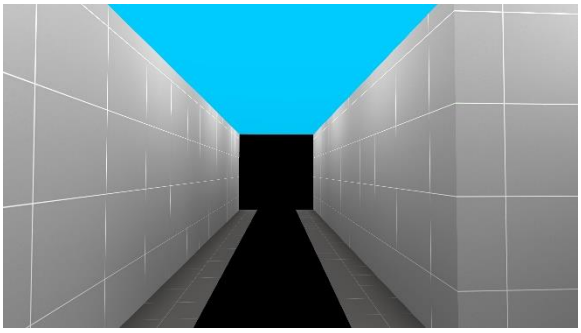
The mazes have a total of 12 turns, divided into 3 equal-length sections comprising of 4 turns. The 3 sections of a maze denote the 3 difficulties of the trial: 4 turns/1 section (easy), 8 turns/2 sections (medium) and 12 turns/3 sections (hard). For each difficulty level, 4 turns are added onto the preceding turns previously navigated. In the first difficulty level, 4 new turns (section 1) are learned in isolation; in the second difficulty level, 4 new turns are added onto the 4 turns from section 1, totalling 8 turns; and in the third difficulty level, 4 new turns are added onto the already learned 8 turns from section 1 and 2, totalling 12 turns. End squares placed after the 4^{th}, 8^{th} and 12^{th} junction were only active during the corresponding difficulty level.

Mazes also varied depending on the condition. During learning and guided phases, a black line was visible on the floor showing participants the correct path through the maze (Figure 5.3a), and lead towards a black wall denoting the end of the maze section (Figure 5.3b). During recall phases, there was no black path nor black wall to show the end of the maze (Figure 5.3c). A fixation cross appeared and disabled locomotion for 5 seconds prior to each run (Figure 5.3d).

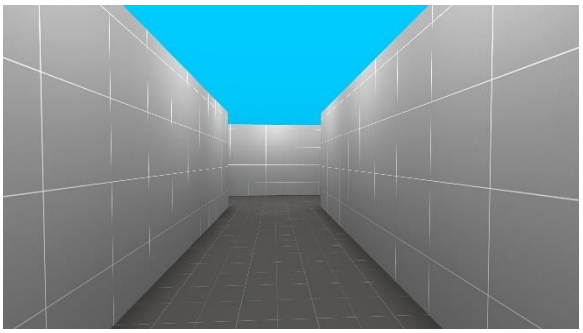
(a) View from a junction in learning or guided trial



(b) View of the end of a maze in a learning or guided trial



(c) View from a corridor in a recall trial



(d) View of a fixation cross in a recall trial

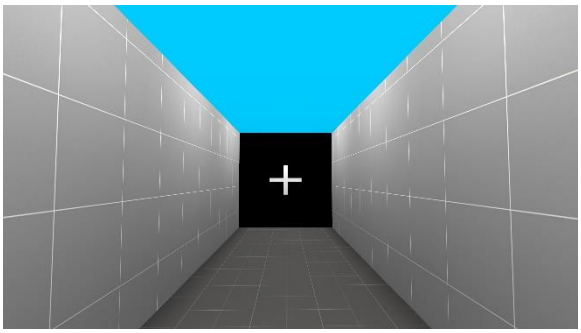


Figure 5.3: A series of screenshots from the maze in difference locations and conditions. The black line showing the correct navigation of a maze route at a junction is during the learning and guided trials is shown in (a). The black wall shown end of the current guided route is shown in (b). An example of approaching a junction in a recall trial is seen in (c). (c) also shows what would be visible to a participant after turning 90 degrees at a junction for both the correct and wrong paths. Moreover, (c) also represents the end point of recall trials, where no marker of the end is provided. The fixation cross before presented before every maze run is presented in (d).

Each maze section contains the same number of squares to navigate for consistency between sections and mazes. No directional information was provided in the maze other than the path presented in the leaning and guided phases. The floor was dark grey with a white grid pattern, and the walls were a lighter grey with a white grid pattern. The grid pattern was carefully sized and aligned between the floor and wall to prevent any minor misalignment providing a clue as to the correct direction. Similarly, the ‘sky’ was a solid blue colour, with no clouds or other visual markings which could introduce directional hints or distractions. Each grid-square within the maze had a single light source placed directly into its centre, so the maze was consistently and evenly lighted throughout.

5.2.2.4) Maze Criteria

The mazes constructed were subject to the following four rules to ensure mazes were distinct but balanced in complexity:

1. The maze must have a clear path from start to end without any overlapping squares.
2. Repeated patterns of 2 turns comprising of 50% or more of the total maze, and patterns of 3, 4 or 6 of over 50% of the maze are not allowed.
    a. Patterns of 2 turns that total of 6 or more (50% of the maze)
    b. Patterns of 3 turns that comprise of 9 or more turns (75% of the maze)
    c. Patterns of 4 turns that comprise of 8 or more turns (66% of the maze)
    d. Patterns of 6 turns repeated twice (100% of the maze)
3. The same turn-sequence section cannot be used in any other maze regardless of position.
4. The final mazes cannot be ‘mirrored’ or ‘reversed’ from each other.

Ten potential turn sequences (Table 5.1) were identified and placed into every unique combination of 3 sections without repetition, totalling 720 potential mazes. Potential mazes that included sets of 3 of the same turns sequentially, which would result in overlap, were removed to leave 348 potential mazes. From here, the first 36 potential mazes which do not contain two turn sequences in the same positions were extracted. Any maze after the first in a list that had the same turn-sections in the first and second phase position, the second and third phase position, or the first and third phase position were removed. The pool of 36 candidate mazes then had a random number assigned to them, and were reordered into a list based on this number from lowest to highest. The final mazes were examined individually until two appropriate mazes could be constructed based on the first and second criteria. It was found that, despite preventing mazes containing three sequential left or right turns, sometimes the turn sequences would result in overlap when balancing the length of the corridors, and thus were unsuitable. After selection of the first maze, the third and fourth criteria were also applied. All the mazes considered and tested, including the two primary mazes selected, are shown in Table 5.2.

Table 5.1: *All possible turn sequence sections applicable to the maze parameters. Left turns in the sequence are marked as 'L', and right turns are marked as 'R'.*

A	LLRL	**F**	RLLR
B	LLRR	**G**	RLRL
C	LRLL	**H**	RLRR
D	LRLR	**I**	RRLL
E	LRRL	**J**	RRLR

Table 5.2: *The outcome of the maze testing and selection process for the primary mazes. Left turns in the sequence are marked as 'L', and right turns are marked as 'R'. Accepted mazes are highlighted in green. The criteria which rejected mazes which included sections previously being used in accepted mazes was only applied after the first maze was accepted.*

Maze ID	Phase 1	Phase 2	Phase 3	Full Sequence	Accept/Reject
12	F: RLLR	A: LLRL	C: LRLL	RLLRLLRLLRLL	Rejected, RLL x4/Overlap
25	F: RLLR	B: LLRR	D: LRLR	RLLRLLRRLRLR	Accepted, Maze-A
31	A: LLRL	D: LRLR	E: LRRL	LLRLLRLRLRRL	Rejected, LR x3
16	J: RRLR	G: RLRL	E: LRRL	RRLRRLRLLRRL	Rejected, RRL x3/Overlap
13	I: RRLL	H: RLRR	C: LRLL	RRLLRLRRLRLL	Rejected, Overlap
28	C: LRLL	G: RLRL	H: RLRR	LRLLRLRLRLRR	Rejected, LR x3
2	F: RLLR	D: LRLR	A: LLRL	RLLRLRLRLLRL	Rejected, F used
26	E: LRRL	I: RRLL	G: RLRL	LRRLRRLLRLRL	Rejected, reverse of F-B-D
33	D: LRLR	B: LLRR	A: LLRL	LRLRLLRRLLRL	Rejected, B Used
11	A: LLRL	J: RRLR	G: RLRL	LLRLRRLRRLRL	Rejected, RLR x3
36	A: LLRL	I: RRLL	H: RLRR	LLRLRRLLRLRR	Rejected, LLRLRRx2
5	J: RRLR	H: RLRR	B: LLRR	RRLRRLRRLLRR	Rejected, RRL x3
32	G: RLRL	I: RRLL	F: RLLR	RLRLRRLLRLLR	Rejected, F-B-D in reverse
35	C: LRLL	I: RRLL	J: RRLR	LRLLRRLLRRLR	Rejected, LLRR x2
17	G: RLRL	C: LRLL	I: RRLL	RLRLLRLLRRLL	Rejected, RLL x3
24	C: LRLL	H: RLRR	D: LRLR	LRLLRLRRLRLR	Rejected, D used
30	D: LRLR	E: LRRL	G: RLRL	LRLRLRRLRLRL	Rejected, D used
34	H: RLRR	D: LRLR	B: LLRR	RLRRLRLRLLRR	Rejected, D used
29	B: LLRR	E: LRRL	C: LRLL	LLRRLRRLLRLL	Rejected, B used
21	J: RRLR	C: LRLL	G: RLRL	RRLRLRLLRLRL	Rejected, 4x LR
19	G: RLRL	F: RLLR	H: RLRR	RLRLRLLRRLRR	Rejected, 4x RL
4	D: LRLR	A: LLRL	I: RRLL	LRLRLLRLRRLL	Rejected, D used
15	I: RRLL	J: RRLR	F: RLLR	RRLLRRLRRLLR	Rejected, RRLLRR
1	H: RLRR	E: LRRL	F: RLLR	LRRLRLLRRLLR	Rejected, F used
3	E: LRRL	J: RRLR	B: LLRR	LRRLRRLRLLRR	Rejected, B used
20	B: LLRR	A: LLRL	G: RLRL	LLRRLLRLRLRL	Rejected, B used
14	F: RLLR	C: LRLL	J: RRLR	RLLRLRLLRRLR	Rejected, F used
22	E: LRRL	C: LRLL	H: RLRR	LRRLLRLLRLRR	Rejected, Overlap
23	B: LLRR	C: LRLL	F: RLLR	LLRRLRLLRLLR	Rejected, B used
8	C: LRLL	J: RRLR	A: LLRL	LRLLRRLRLLRL	Accepted, Maze-B

7	G: RLRL	J: RRLR	E: LRRL	RLRLRRLRLRRL	Rejected, 4x RL
10	A: LLRL	E: LRRL	D: LRLR	LLRLLRRLLRLR	Rejected, D used
6	A: LLRL	G: RLRL	J: RRLR	LLRLRLRLRRLR	Rejected, 4x LR
27	D: LRLR	F: RLLR	E: LRRL	LRLRRLLRLRRL	Rejected, D used
9	E: LRRL	F: RLLR	D: LRLR	LRRLRLLRLRLR	Rejected, F used
18	H: RLRR	A: LLRL	D: LRLR	RLRRLLRLLRLR	Rejected, D used

A third maze was designed as a backup in case a dataset was lost due to technical issue, but the participant otherwise completed the maze. The maze was constructed from the remaining 4 unused turn sequences (E, G, H and I, Table 5.1). Each potential combination of the 4 turn sequences was tested. Of the 24 potential combinations, 13 mazes had no triple-repeated-turns, were randomly ordered, and tested until an appropriate maze which fitted the criteria was found (Table 5.3).

Table 5.3: *The outcome of the maze testing and selection process for the 'backup' maze. Left turns in the sequence are marked as 'L', and right turns are marked as 'R'. Accepted mazes are highlighted in green. After a suitable maze was found, no additional testing was performed.*

Maze ID	Phase 1	Phase 2	Phase 3	Full Sequence	Accept/Reject
2-6	G: RLRL	E: LRRL	I: RRLL	RLRLLRRLRRLL	Accepted, Maze-C
2-2	E: LRRL	G: RLRL	I: RRLL	LRRLRLRLRRLL	N/A
2-5	G: RLRL	E: LRRL	H: RLRR	RLRLLRRLRLRR	N/A
2-8	G: RLRL	I: RRLL	H: RLRR	RLRLRRLLRLRR	N/A
2-7	G: RLRL	H: RLRR	E: LRRL	RLRLRLRRLRRL	N/A
2-10	H: RLRR	E: LRRL	I: RRLL	RLRRLRRLRRLL	N/A
2-9	H: RLRR	E: LRRL	G: RLRL	RLRRLRRLRLRL	N/A
2-3	E: LRRL	I: RRLL	G: RLRL	LRRLRRLLRLRL	N/A
2-11	I: RRLL	G: RLRL	H: RLRR	RRLLRLRLRLRR	N/A
2-13	I: RRLL	H: RLRR	E: LRRL	RRLLRLRRLRRL	N/A
2-1	E: LRRL	G: RLRL	H: RLRR	LRRLRLRLRLRR	N/A
2-4	E: LRRL	I: RRLL	H: RLRR	LRRLRRLLRLRR	N/A
2-12	I: RRLL	G: RLRL	E: LRRL	RRLLRLRLLRRL	N/A

The final 3 maze paths are as follows:

- Maze-A: RIGHT-LEFT-LEFT-RIGHT-LEFT-LEFT-RIGHT-RIGHT-LEFT-RIGHT-LEFT-RIGHT
- Maze-B: LEFT-RIGHT-LEFT-LEFT-RIGHT-RIGHT-LEFT-RIGHT-LEFT-LEFT-RIGHT-LEFT
- Maze-C: RIGHT-LEFT-RIGHT-LEFT-LEFT-RIGHT-RIGHT-LEFT-RIGHT-RIGHT-LEFT-LEFT

To prevent overlap between maze squares, the length of ‘corridors’ varied between 3 and 5 grid-squares. This does mean the differences in length of the corridors before each turn could be noticed by a particularly astute participant; however the visual similarity, relatively quick speed of navigation, and lack of any pattern between corridor length and turn direction means it is unlikely a participant will utilise this information in learning the turns.

Each trial was separated by a ‘preparation room’ (Figure 5.4), which contained two buttons on opposing walls. These served as rooms to explain the next trial to the participant, and to apply or remove the HMD-VR as required. The button could only be interacted with by the researcher’s keyboard. Moreover, a short practice course was produced for the familiarisation procedure (Figure 5.5). Unlike the maze which had wrong turns and walls past the eyeline, the practice course had a singular correct path leading to a black box at the end which served as the endpoint. The path contained 90 degree and 180 degree turns to emulate the movement required for correct and incorrect turns respectively in the main maze. The walls were halfway between the floor and default eyeline so participants could see the complete path whilst learning to use the controller or the HMD-VR device.

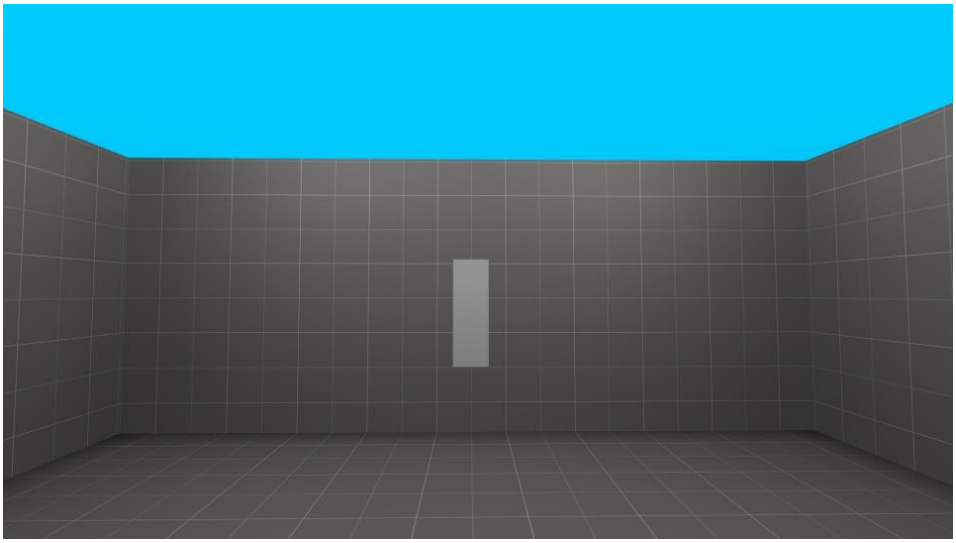


Figure 5.4: *Screenshot of the inside of a 'preparation room'. The view from the opposite rotation is identical.*

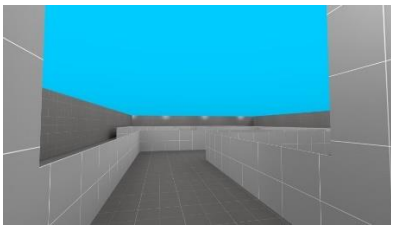
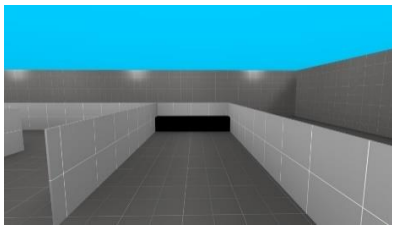
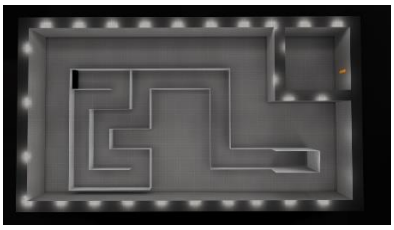
(a) View from the start of the familiarisation path	(b) View of the end of the familiarisation path	(c) Top-down view of the familiarisation path
		

Figure 5.5: Screenshots of the pathway used for the familiarisation period. A screenshot from the start is shown in (a), showing the heightened walls of the start room and the lower walls of the path. The end point is shown in (b). A top down view of the familiarisation path and a preparation room in the top right is shown in (c).

As Garry's Mod is built upon a first-person shooter game, weapons are provided to the player by default and 'heads-up display' elements such as health and ammo counter are present on the screen. Any such 'video game' element are removed or disabled as the participant is placed into the VE, as leaving the participant armed may lead to extraneous load through distraction, or attempts to 'cheat' by marking the correct path using bullet holes in the maze.

Movement speed between displays was verified by comparing the time taken to navigate three straight-line courses comprised of the same number of squares of

each maze difficulty level. The default DB-VR speed was reduced from an arbitrary value of 200 to 177, which resulted in a >0.1s difference between displays for path lengths of each difficulty level (Table 5.4). Similarly, the horizontal sensitivity (turn speed) was also balanced between conditions. The DB-VR was slowed using the horizontal speed slider in the settings (Figure 5.6) until the time taken to complete a single 180-degree turn was ~1s in both conditions.

Table 5.4: *The time taken to complete straight line paths totalling the same distance as the 3 lengths of the turn sequences between display types.*

	DB-VR (177)	HMD-VR (200)	Difference
4-Turn Length Average	14.42s	14.44s	0.02s
8-Turn Length Average	28.87s	28.95s	0.08s
12-Turn Length Average	43.34s	43.42s	0.08s

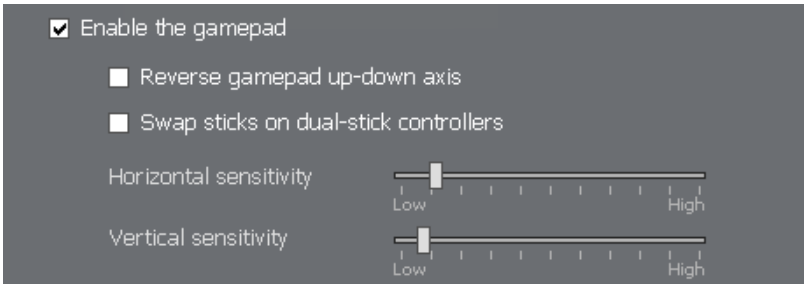


Figure 5.6: *A screenshot of the Garry's mod settings used for the horizontal turn speed in the DB-VR condition.*

The virtual avatar given to participants was made invisible for three reasons. Firstly, it has previously been found that utilising virtual avatars can have negative effects on performance [146]. Secondly, as the participant was seated, there may be a visual-vestibular disconnect in seeing legs walking if the participant looks down. Thirdly, the software utilised to enable controller support in VR places the virtual hands in the participants' field of vision. Not only could this be a distracting element, but the hands could not be placed into the same position in the DB-VR condition.

5.2.2.5) Behavioural Measures

A preliminary questionnaire (Appendix 7) was used to capture information about each participant's experience with HMD-VR, video games, and gamepad controllers. Handedness was also captured in the preliminary questionnaire using a modified Edinburgh Handedness Inventory [250], with additional measures of prior experience with virtual reality, prior experience with video games, and experience with video game input methods. The 16-item SSQ was applied at 3 points: prior to the commencement of the input and HMD-VR familiarisation procedure, following the completion of the DB-VR block, and following the completion of the HMD-VR block. The SSQ was read aloud by the researcher and completed verbally by the participant for consistency between the conditions.

Within the maze task, three behavioural measures were recorded per trial: completion time; number of wrong turns; and if the trial was completed. Completion time was recorded from when the participant was able to move to when the contact was made with the end square. Wrong turns were counted each time the 'wrong path' block was entered following a junction, and the total number of wrong turns per maze run were recorded. Recall trials were considered successfully completed if the end square was reached without any wrong turns being taken. For the learning and guided phases, the participant had to reach the end square and was considered successful regardless of wrong turns. A 5-minute time limit was implemented for each condition block, which if reached would count the current run as failed. Each trial comprised of between 1-6 runs of the maze with a minimum of 1 complete run being required for inclusion in the statistical analysis.

5.2.3) Experimental Design

The current experiment uses a 2x3x3 within-subjects design to examine the effect of display method on spatial navigation performance across three navigation types and three maze difficulties. Behavioural and EEG measures of WML are compared between the two display methods of HMD-VR and DB-VR, three navigation types of learning, recall and guided navigation, and three difficulty levels of 4-turn, 8-turn and 12 turn mazes. The first two mazes constructed used for the task are also

behaviourally compared in a 2x3x3 design to identify differences in maze difficulty that may influence experimental outcomes. Behavioural measures are compared between Maze-A and Maze-B, the three navigation types, and the three maze difficulties.

5.2.4) Experimental Procedure

Upon arriving at the laboratory, participants were provided with a participant information sheet (Appendix 3), consent form (Appendix 4), a health questionnaire (Appendix 5), and the preliminary questionnaire (Appendix 10) to be completed in this order. Following completion of these forms, the 'baseline' SSQ is also recorded. The maze task was completed over the familiarisation, the DB-VR and the HMD-VR phases. The familiarisation phase was always completed first, whilst the order of the DB-VR and HMD-VR phases counterbalanced between participants.

The familiarisation phase introduces participants to the input method used to navigate the maze, and the use of HMD-VR during navigation. Participants were instructed to navigate a training path as many times as possible within a 5-minute time limit, starting slowly and gradually increasing movement speed. Each run of the practice path consists of a 5-second fixation cross, for which locomotion is disabled, and ends with an endpoint that restarts the process. Upon reaching the 5-minute limit, the participant was teleported back to the preparation room. The training path was then navigated again whilst the participant was in HMD-VR. Upon completion of the familiarisation phase, the EEG headcap was fitted to the participant.

Both the DB-VR and HMD-VR phases follow the same 10-step procedure comprising 4 types of trial blocks: eyes open; maze learning trials; maze recall trials; and maze guided trials. The maze trial blocks were completed 3 times (once per difficulty level) for each of the 2 display conditions. For the learning phase, participants were instructed to follow a path from the start to the end of the maze, and attempt to learn the correct sequence of turns. Upon reaching the end square, the participant was teleported back to the start of the maze if less than three successful maze runs had

been conducted, or to the preparation room if either three successful maze runs or six total maze runs had been completed.

Each learning phase is followed by a recall phase, where participants are instructed to navigate the same maze unguided. The recall phase is completed when the maze is navigated a total of 3 times without error, or 6 attempted runs including failed runs had been completed. A limit of 6 attempts was selected to prevent participants having more opportunity to successfully complete the lower difficulty conditions, as 12 turn mazes took ~45 seconds to complete (without mistakes), whereas the 4 and 8 turns took ~15 and ~30 seconds respectively. A total of 6 attempts of the 12 turn condition would therefore take at least 4 minutes and 30 seconds, leaving insufficient time to complete a seventh run. The limit also acted to prevent increased frustration or similar reactions at repeated failures which may result in a participant ceasing attempts [401]. Participants were instructed that in the event they make a wrong turn (denoted by a dead end), they should proceed to the end of the maze to restart. Upon reaching the end of the maze, participants are teleported to either the start of the maze if less than 3 correct or 6 total runs had been completed, or back to the preparation room should these criteria be met. Upon completion of a recall phase, participants were asked to rate the subjective difficulty of the maze on a scale of 1-10 (1 = very easy, 10 = very hard).

Upon completion of 3 learning and recall phases, the participants underwent three ‘guided’ phases. The guided phase is functionally identical to the learning phase, but with instruction to follow the path without memorising the turn sequence. Upon completion of the third guided phase in either display condition, the SSQ is recorded, and the display block is completed.

Upon completion of the second display block, participants were asked if they had a method or strategy of remembering the turn sequences in the maze, before having any EEG and HMD-VR equipment removed, provided a debrief form, and thanked

for their time. Participants were given the opportunity to ask any final questions, before leaving the laboratory.

The HMD-VR was calibrated for visual clarity each time it is placed onto the participant, and verified by having participants read aloud a passage visible on the wall within the VE explaining the maze task. Before each block, participants are placed into a virtual preparation room, where instructions for the following trial are verbally provided. The researcher was positioned in the laboratory facing away from the computer screen and out of view of the participant. Participants were therefore instructed to inform the researcher when they completed a trial block.

In instances where EEG system-related technical issues were encountered which prevented data collection for one display condition, and providing the other condition was completed successfully, participants were invited back to the laboratory. In such instances, the participant completed the third maze designed exclusively for this occurrence. In instances when this occurred, the participant underwent a familiarisation period as normal if the HMD-VR data was being re-captured, or only the DB-VR familiarisation period if the DB-VR data is being recaptured. The EEG data acquired was then included with the main analysis.

After the loss of several early datasets due to technical problems with the EEG device corrupting recorded data, the procedure was modified to include time to reset the EEG recording during each preparation room. For the first 11 participants, 2 datasets were recorded for the experiment, one for each display condition. However, it was found that the EEG recording device sometimes corrupted data during some point in the recording process which was undetectable until later analysis. Therefore, from the 12th participant onwards separate datasets were recorded for each trial block, leading to 20 recordings in total. The restarting of the EEG recording system did extend the time spent in the preparation room by ~30s each time it was visited, but this was deemed necessary to prevent usable data being lost or comparisons between data being excluded from the later ANOVA analysis.

5.2.5) Electroencephalography Data Preprocessing

The EEG data was preprocessed using the EEGLAB toolbox [261] for MATLAB [262] and the ERPLAB plugin [263]. Data processing was conducted over three stages (Figure 5.7).

In the first phase, the data was converted from a biowav (.bwv) file type to the EEG dataset (.set) format for EEGLAB compatibility. Electrode location data was applied to the dataset, followed by the application of a 0.5Hz highpass and a 50Hz notch filter. Datasets with separate recordings for each phase were appended to match the datasets with continuous recordings for each display. All datasets had a temporary 40Hz notch filter applied to identify noisy or disconnected channels, which were then interpolated on the non-40Hz filtered data.

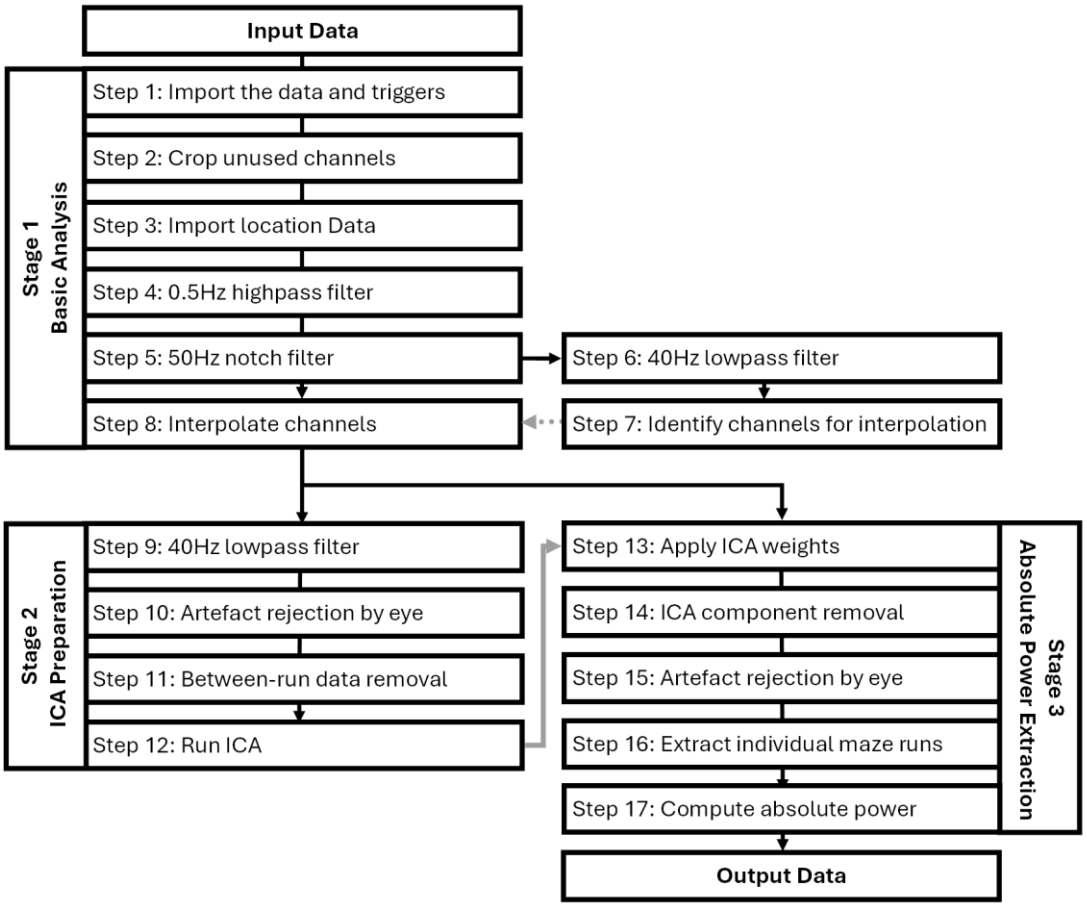


Figure 5.7: *EEG analysis chart showing all 3 stages of the absolute power extraction. Solid black lines show where the entire dataset is carried through to the next processing step. The dotted grey line connecting the 'Identify Channels to Interpolate' to 'Interpolate channels' in stage 1 represents information from the former stage being used in the latter, but no actual data. The grey line collecting 'Run the ICA' in stage 2 to 'Apply the ICA Weights' in stage 3 represents the ICA weights calculated only being applied to the latter.*

The second phase prepared the data for ICA to identify artifacts originating from the eyes. The continuous data had a 40Hz lowpass filter applied, and artifact rejection for non-eyeblink artifacts was performed on the continuous data. Recorded EEG data between each trial block was removed, and the ICA weightings were acquired.

The third phase prepared and extracted the absolute power values of the theta and alpha frequency bands. The dataset produced in stage 1 has the ICA weightings from stage 2 applied, and eye artifact components removed. Data between trial blocks was also removed and manual artifact rejection was performed on the

continuous data. Finally, each individual run of the maze was extracted from the dataset, and converted to power spectral density using 1 second windows with 50% overlap. The absolute power of mean theta activity (5-8Hz) in the Fz electrode and mean alpha activity (10-13Hz) in the Pz electrode was extracted for each trial block across conditions. The TAR is calculated by dividing the absolute powers of the theta activity in the Fz electrode by the absolute power of the alpha activity in the Pz electrode [338].

5.2.6) Data and Statistical Analysis

Statistical analysis was performed using SPSS 28 [264] and Microsoft Excel [265]. Behavioural data was compared between displays (DB-VR and HMD-VR) and mazes (Maze-A and Maze-B). The completion time was calculated from the end of the baseline period to the end point of a maze being reached. Total completion time is analysed over two 2x2x3 ANOVA, one comparing between displays (display [DB-VR, HMD-VR] x condition [Learning, Recall] x difficulty level [4-turns, 8-turns, 12-turns]) and the other between mazes (2x2x3 design (maze [Maze-A, Maze-B] x condition [Learning, Recall] x difficulty level [4-turns, 8-turns, 12-turns]). A second set of completion time ANOVA comparisons using the same factors was conducted for the time to navigate the final section (comprising of the ‘new’ 4 turns and corridors added at that difficulty level) of each completed maze run. The completion time for the final 4 turns was extracted as they comprise of a consistent distance between difficulty phases and mazes, and thus allow a comparison of WML between conditions. The average number of wrong turns per recall trial was calculated by dividing the total number of wrong turns by the total number of runs per recall block. Both the average number of wrong turns and the average subjective difficulty rating for each recall block was compared using a repeated measures 2x3 ANOVA between displays (DB-VR, HMD-VR x difficulty level (4-turns, 8-turns, 12-turns), and using a repeated measures 2x3 ANOVA between mazes (maze [Maze-A, Maze-B] x difficulty level [4-turns, 8-turns, 12-turns]). Statistical significance was determined by a p-value of less than $p<0.05$.

The absolute theta power, absolute alpha power, and the TAR data for the correct run of each trial block was analysed using a series of 3-factor repeated measures design ANOVA using a 2x3x3 design (display [DB-VR, HMD-VR] x condition [Learning, Recall, Guided] x difficulty level [4-turns, 8-turns, 12-turns]) to compare differences between conditions.

All ANOVAs conducted used Mauchly's test of sphericity to identify main effects and interactions which violated the sphericity assumption. When a violation was found, the degrees-of-freedom of the main effect or interaction was adjusted using the Greenhouse-Geisser correction when the Greenhouse-Geisser Epsilon was under 0.75, and the Huynh-Feldt correction when the Greenhouse-Geisser Epsilon was over 0.75. Bonferroni-adjusted pairwise comparisons were conducted for post-hoc analysis of significant main effects and interactions.

5.3) Results

5.3.1) Behavioural Results

5.3.1.1) Completion Time Between Display Conditions

5.3.1.1.1) Total Maze Navigation Time

A 2x2x3 repeated measures ANOVA (display [DB-VR, HMD-VR] x condition [Learning, Recall] x difficulty level [4-turns, 8-turns, 12-turns]) was conducted for the total completion time of successfully completed maze runs between displays. Mauchly's Test of Sphericity was violated by the interactions between display x difficulty level ($\chi^2(2)=6.95$, p0.031, $\varepsilon=0.7$), which was adjusted using the Greenhouse-Geisser correction. The mean completion time between displays, conditions and difficulties is displayed in Table 5.5.

Table 5.5: *The mean and SEM of the total maze completion time for the learning and recall trials presented in DB-VR and HMD-VR across the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

	Display	**DB-VR**			**HMD-VR**		
	Difficulty	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Total Maze Completion Time (s)	**Learning**	15 ± 0.7	28.89 ± 1.07	41.64 ± 0.69	14.36 ± 0.64	27.76 ± 0.78	40.62 ± 0.94
	Recall	14.38 ± 0.63	28.53 ± 1.33	42.09 ± 0.93	14.17 ± 0.63	26.89 ± 0.5	41.87 ± 1.01

The main effects of difficulty level reached significance (Table 5.6). Post hoc comparisons found between conditions, 4-turn mazes (M=14.48s ± 0.47) were completed significantly faster than 8-turn maze trials (M=28.02 ± 0.78) (t(13)=-32.29, p=<0.001, d=-8.63) and 12=turn maze trials (M=41.55 ± 0.63) (t(13)=-68.77, p=<0.001, d=-18.38). The 8-turn maze trials were also completed significantly faster than the 12-turn maze trials (t(13)=-35.78, p=<0.001, d=-9.56).

Table 5.6: *Main effects and interactions of the total maze completion time ANOVA between displays, navigation conditions and difficulties. n=14.*

Within Subjects Effect	**F**	**df**	**df (error)**	**p**	**η_p^2**
Display	1.23	1	13	0.29	0.09
Condition	0.03	1	13	0.87	0.00
Difficulty Level	2320.03	2	26	**<0.001***	0.99
Display x Condition	0.15	1	13	0.71	0.01
Display x Difficulty Level	0.96	1.39	18.06	0.37	0.07
Condition x Difficulty Level	3.21	2	26	0.06	0.20
Display x Condition x Difficulty Level	0.60	2	26	0.56	0.04

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

5.3.1.1.2) Final Section Navigation Time

A second 2x2x3 repeated measures ANOVA (display [DB-VR, HMD-VR] x condition [Learning, Recall] x difficulty level [4-turns, 8-turns, 12-turns]) was conducted for the completion time of the final section only for successfully completed maze runs between displays. Mauchly's Test of Sphericity was not violated by any interaction or main effect. The mean completion time between displays, conditions and difficulties is displayed in Table 5.7.

Table 5.7: *The mean and SEM of the final section completion time for the learning and recall trials presented in DB-VR and HMD-VR across the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

	Display	**DB-VR**			**HMD-VR**		
	Difficulty	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Total Maze Completion Time (s)	**Learning**	15 ± 0.7	13.81 ± 0.74	13.07 ± 0.21	14.36 ± 0.64	13.03 ± 0.39	12.62 ± 0.26
	Recall	14.38 ± 0.63	13.57 ± 0.78	13.27 ± 0.42	14.17 ± 0.63	12.43 ± 0.14	12.97 ± 0.4

The main effects of difficulty level reached significance (Table 5.8). Post hoc comparisons found between conditions, 4-turn mazes (M=14.48s ± 0.47) were completed significantly slower than 8-turn maze trials (M=13.21 ± 0.4) ($t(13)=-4.47$, $p=<0.001$, $d=1.19$) and 12=turn maze trials (M=12.98 ± 0.2) ($t(13)=3.846$, $p=0.002$, $d=1.03$).

Table 5.8: *Main effects and interactions of the final section maze completion time ANOVA between displays, navigation conditions and difficulties. n=14.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Display	1.77	1	13	0.21	0.12
Condition	0.58	1	13	0.46	0.04
Difficulty Level	12.56	2	26	**<0.001***	0.49
Display x Condition	0.02	1	13	0.90	0.00
Display x Difficulty Level	0.70	2	26	0.51	0.05
Condition x Difficulty Level	1.22	2	26	0.31	0.09
Display x Condition x Difficulty Level	0.38	2	26	0.69	0.03

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

5.3.1.2) Subjective Difficulty Rating Between Display Conditions

A 2x3 repeated measures ANOVA (display [DB-VR, HMD-VR] x difficulty level [4-turns, 8-turns, 12 turns]) was conducted for the subjective difficulty level rating between displays. The mean subjective difficulty rating and SEM between displays can be seen in table 5.9, and the ANOVA is summarised in table 5.10. The main effect of difficulty level ($\chi^2(2)=13.28$, $p\leq0.001$, $\varepsilon=0.6$) violated Mauchly's Test of Sphericity, and was adjusted using the Greenhouse-Geisser correction. Only the main effect of difficulty level reached significance (Table 5.10), with post hoc tests finding that the 4-turn mazes (M=1.5 ± 0.2) were rated significantly lower on the measure of subjective difficulty than the 8-turn mazes (M=3.2 ± 0.3) ($t(13)=-1.71$, $p\leq0.001$, $d=-1.72$) and 12-turn mazes (M=5 ± 0.5) ($t(13)=-3.54$, $p\leq0.001$, $d=-1.97$). The difference between 8-turn and 12-turn mazes also reached significance, with 8-turn mazes ($t(13)=-1.82$, $p\leq0.001$, $d=-1.74$) being rated as a lower subjective difficulty.

Table 5.9: The mean and SEM of the subjective difficulty ratings on a scale of 1 to 10 for the recall trials presented in DB-VR and HMD-VR across the 4-turn, 8-turn and 12-turn maze difficulties. n=14.

Display	**DB-VR**			**HMD-VR**		
Difficulty level	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Subjective Difficulty Rating	1.5 ± 0.2	3.36 ± 0.37	5.07 ± 0.52	1.43 ± 0.17	3 ± 0.3	4.93 ± 0.54

Table 5.10: *Main effects and interactions of the subjective difficulty rating on a scale of 1 to 10 ANOVA between displays and difficulties. n=14.*

Within Subjects Effect	**F**	**df**	**df (error)**	**p**	**η_p^2**
Display	0.883	1	14	0.363	0.059
Difficulty level	48.645	1.185	16.588	**<0.001***	0.777
Display x Difficulty level	0.528	2	28	0.595	0.036

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

5.3.1.3) Average Number of Wrong Turns Between Display Conditions

A 2x3 repeated measures ANOVA (display [DB-VR, HMD-VR] x difficulty level [4-turns, 8-turns, 12 turns]) was conducted for the average number of wrong turns between displays. The mean number of wrong turns per run and SEM between displays can be seen in table 5.11, and the ANOVA is summarised in table 5.12. Mauchly's Test of Sphericity was violated by the main effect of difficulty level ($\chi^2(2)=35.5$, $p\leq0.001$, $\varepsilon=0.51$)) and interaction between Display x Diff ($\chi^2(2)=12.38$, $p=0$, $\varepsilon=0.61$), both of which was adjusted using the Greenhouse-Geisser correction. No main effects or interactions reached significance (table 5.12).

Table 5.11: *The mean and SEM of the number of wrong turns per run for the recall trials presented in DB-VR and HMD-VR across the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

Display	**DB-VR**			**HMD-VR**		
Difficulty level	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Average Number of Wrong Turns	0.07 ± 0.06	0.04 ± 0.02	0.5 ± 0.31	0.05 ± 0.05	0.02 ± 0.02	0.31 ± 0.2

Table 5.12: *Main effects and interactions of the average number of wrong turns ANOVA tests between displays and difficulties. n=14.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Display	1.25	1	13	0.28	0.09
Difficulty level	2.11	1.03	13.35	0.17	0.14
Display x Difficulty level	0.58	1.22	15.82	0.49	0.04

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

5.3.1.4) Completion Time Between Mazes

5.3.1.4.1) Total Maze Navigation Time

The total completion time for whole maze runs between mazeIDs was compared using a 2x2x3 repeated measures ANOVA (mazeID [Maze-A, Maze-B] x condition [Learning, Recall] x difficulty level [4-turns, 8-turns, 12-turns]). The mean completion times and SEM between mazes can be seen in table 5.13. Mauchly's Test of Sphericity was violated by the interaction between mazeID x difficulty level ($\chi^2(2)=7.64$, p0.02, $\varepsilon=0.68$), which was adjusted using the Greenhouse-Geisser correction.

Table 5.13: *The mean and SEM of the total completed maze navigation time for the learning and recall trials presented in Maze-A and Maze-B across the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

	MazeID	**Maze-A**			**Maze-B**		
	Difficulty	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Adjusted Maze Completion Time (s)	Learning	13.93 ± 0.26	27.35 ± 0.46	40.37 ± 0.51	15.43 ± 0.87	29.29 ± 1.2	41.89 ± 1.02
	Recall	14.13 ± 0.61	26.84 ± 0.38	40.87 ± 0.6	14.42 ± 0.65	28.59 ± 1.37	43.09 ± 1.15

The main effects of mazeID and difficulty level reach significance (Table 5.14). Post Hoc comparisons between mazeIDs find that Maze-A (M=27.25s ± 0.36) was completed significantly faster than Maze-B (M=28.79s ± 0.91) (t(13)=-2.422,

p=0.031, d=0.44). Between difficulty levels, post hoc comparisons find that the 4-turn condition (M=14.48s ± 0.47) was completed significantly faster than the 8-turn (M=28.08s ± 0.78) (t(13)=-13.54, p=<0.001, d=-8.38) and 12-turn conditions (M=51.55s ± 0.68) (t(13)=-27.08, p=<0.001, d=-18.38). The 8-turn condition was also completed significantly faster than the 12-turn condition (t(13)=-13.54, p=<0.001, d=-9.56).

Table 5.14: *Main effects and interactions of the total maze completion time ANOVA between mazeIDs, navigation conditions and difficulties. n=14.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Maze ID	5.86	1	13	**0.03***	0.31
Condition	0.03	1	13	0.87	0.00
Difficulty Level	2320.03	2	26	**<0.001***	0.99
MazeID x Condition	0.14	1	13	0.72	0.01
MazeID x Difficulty Level	1.18	1.36	17.68	0.31	0.08
Condition x Difficulty Level	3.21	2	26	0.06	0.20
MazeID x Condition x Difficulty Level	1.27	2	26	0.30	0.09

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

5.3.1.4.2) Final Section Navigation Time

The completion time of the final section for completed maze runs between mazeIDs was compared using a 2x2x3 repeated measures ANOVA (mazeID [Maze-A, Maze-B] x condition [Learning, Recall] x difficulty level [4-turns, 8-turns, 12-turns]). The mean completion times and SEM between mazes can be seen in table 5.15. Mauchly's Test of Sphericity was not violated by any main effect or interaction.

Table 5.15: *The mean and SEM of the completion time for the final section of each completed maze for the learning and recall trials presented in Maze-A and Maze-B across the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

	MazeID	**Maze-A**			**Maze-B**		
	Difficulty	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Adjusted Maze Completion Time (s)	Learning	15 ± 0.7	13.81 ± 0.74	13.07 ± 0.21	14.36 ± 0.64	13.03 ± 0.39	12.62 ± 0.26
	Recall	14.38 ± 0.63	13.57 ± 0.78	13.27 ± 0.42	14.17 ± 0.63	12.43 ± 0.14	12.97 ± 0.4

The main effect of difficulty level reached significance (Table 5.16). Post Hoc comparisons between difficulty levels find that the 4-turn condition (M=14.48s ± 0.47) was completed significantly slower than the 8-turn (M=13.21s ± 0.4) (t(13)=4.47, p=<0.001, d=1.17) and 12-turn conditions (M=12.98s ± 0.2) (t(13)=3.85, p=0.002, d=1.03).

Table 5.16: *Main effects and interactions of the final section completion time ANOVA between mazeIDs, navigation conditions and difficulties. n=14.*

Within Subjects Effect	**F**	**df**	**df (error)**	**p**	**η_p^2**
Maze ID	4.48	1	13	0.054	0.26
Condition	0.58	1	13	0.46	0.04
Difficulty Level	12.56	2	26	**<0.001***	0.49
MazeID x Condition	0.94	1	13	0.35	0.07
MazeID x Difficulty Level	0.57	2	26	0.57	0.04
Condition x Difficulty Level	1.22	2	26	0.31	0.09
MazeID x Condition x Difficulty Level	0.83	2	26	0.45	0.06

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

5.3.1.5) Subjective Difficulty Between Mazes

A 2x3 repeated measures ANOVA (mazeID [Maze-A, Maze-B] x difficulty level [4-turns, 8-turns, 12 turns]) was conducted for the subjective difficulty rating between mazeIDs. The mean subjective difficulty rating and SEM between mazeIDs can be seen in table 5.17, and the ANOVA is summarised in table 5.18. It was found that the main effect of difficulty level ($\chi^2(2)=13.28$, $p \leq 0.001$, $\varepsilon=0.6$) violated Mauchly's Test of Sphericity, and was adjusted using the Greenhouse-Geisser correction.

Table 5.17: *The mean and SEM of the subjective difficulty ratings for the recall trials in Maze-A and Maze-B across the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

MazeID	**Maze-A**			**Maze-B**		
Difficulty level	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Subjective Difficulty Rating	1.43 ± 0.2	3.07 ± 0.34	4.64 ± 0.52	1.5 ± 0.17	3.29 ± 0.34	5.36 ± 0.52

Table 5.18: *Main effects and interactions of the subjective difficulty rating ANOVA between MazeIDs and difficulties. n=14.*

Within Subjects Effect	**F**	**df**	**df (error)**	**p**	**η_p^2**
MazeID	2.4	1	13	0.15	0.16
Difficulty level	49.55	1.2	15.58	**<0.001***	0.79
MazeID x Difficulty level	3.84	2	26	**0.04***	0.23

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

The main effect of difficulty level and the interaction between mazeID x difficulty level reached significance. For all post hoc comparisons within the main effect of difficulty level and in the interaction between mazeIDs and difficulty levels, it is found that the 4-turn mazes were rated as a significantly lower subjective difficulty than both the 8-turn and 12-turn mazes; which also significantly differed across comparisons with the 8-turn mazeID being rated lower (Table 5.19). One significant post hoc comparison is found within difficulties and between mazeIDs, with the 12-turn Maze-A being rated as a significantly lower subjective difficulty than the 12-turn Maze-B.

Table 5.19: *Post hoc comparisons of the significant main effects and interactions for the subjective difficulty rating ANOVA between mazeIDs. n=14.*

Comparison	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	P
Difficulty Level	4-Turns	1.46	0.15	8-Turns	3.18	0.31	-6.45	13	-1.71	0.27	**<0.001***
	4-Turns	1.46	0.15	12-Turns	5	0.5	-7.38	13	-3.54	0.48	**<0.001***
	8-Turns	3.18	0.31	12-Turns	5	0.5	-6.5	13	-1.82	0.28	**<0.001***
Difficulty Level x MazeID	4-Turns Maze-A	1.43	0.2	4-Turns Maze-B	1.5	0.17	-0.32	13	-0.07	0.22	0.75
	8-Turns Maze-A	3.07	0.34	8-Turns Maze-B	3.29	0.34	-0.76	13	-0.21	0.28	0.46
	12-Turns Maze-A	4.64	0.52	12-Turns Maze-B	5.36	0.52	-2.69	13	-0.71	0.27	**0.02***
MazeID x Difficulty Level	Maze-A 4-Turns	1.43	0.2	Maze-A 8-Turns	3.07	0.34	-5.68	13	-1.64	0.29	**<0.001***
	Maze-A 4-Turns	1.43	0.2	Maze-A 12-Turns	4.64	0.52	-6.51	13	-3.21	0.49	**<0.001***
	Maze-A 8-Turns	3.07	0.34	Maze-A 12-Turns	4.64	0.52	-5.4	13	-1.57	0.29	**<0.001***
	Maze-B 4-Turns	1.5	0.17	Maze-B 8-Turns	3.29	0.34	-6.36	13	-1.79	0.28	**<0.001***
	Maze-B 4-Turns	1.5	0.17	Maze-B 12-Turns	5.36	0.52	-7.87	13	-3.86	0.49	**<0.001***
	Maze-B 8-Turns	3.29	0.34	Maze-B 12-Turns	5.36	0.52	-6.11	13	-2.07	0.34	**<0.001***

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, $p<0.05$.

5.3.1.6) Wrong Turns Between Mazes

A 2x3 repeated measures ANOVA (maze [Maze-A, Maze-B] x difficulty level [4-turns, 8-turns, 12 turns]) was also conducted for the average number of wrong turns between mazeIDs. The mean number of wrong turns per run and SEM between mazeIDs can be seen in table 5.20, and the ANOVA is summarised in table 5.21. Mauchly's Test of Sphericity was violated by the main effect of difficulty level ($\chi^2(2)=35.5$, $p\leq0.001$, $\varepsilon=0.51$)and the interaction between mazeID x difficulty level ($\chi^2(2)=11.14$, $p=0$, $\varepsilon=0.62$), both of which is adjusted using the Greenhouse-Geisser correction. No main effects or interactions reach significance (Table 5.21).

Table 5.20: *The mean and SEM of the number of wrong turns per run for the recall trials in Maze-A and Maze-B across the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

MazeID	**Maze-A**			**Maze-B**		
Difficulty level	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Average Number of Wrong Turns	1.43 ± 0.2	3.07 ± 0.34	4.64 ± 0.52	1.5 ± 0.17	3.29 ± 0.34	5.36 ± 0.52

Table 5.21: *Main effects and interactions of the average number of wrong turns ANOVA between MazeIDs and difficulties. n=14.*

Within Subjects Effect	**F**	**df**	**df (error)**	**p**	**η_p^2**
MazeID	3.52	1	14	0.08	0.20
Difficulty level	2.21	1.03	14.37	0.16	0.14
MazeID x Difficulty level	1.95	1.24	17.31	0.18	0.12

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

5.3.1.7) Simulator Sickness Questionnaire Results

Separate 1 factor 3-way (Condition [Baseline, DB-VR, HMD-VR]) repeated measures ANOVA analyses were conducted for the four subscales of the SSQ (Total score, Combined Nausea, Oculomotor Disturbance, Disorientation). The mean value and SEM for each SSQ subscale can be seen in table 5.22 and are summarised in table 5.22. Mauchly's Test of Sphericity was violated for the subscale of Disorientation ($\chi^2(2)=18.55$, $p\leq0.001$, $\varepsilon=0.56$), which was adjusted using the Greenhouse-Geisser correction.

Table 5.22: *The mean and SEM of the 4 SSQ measures taken for each display type. n=14.*

Comparison	Baseline	Desktop	HMD-VR
Total Score	5.34 ± 2.24	12.02 ± 3.42	24.84 ± 6.21
Combined Nausea	4.77 ± 2.18	8.18 ± 2.8	21.12 ± 5.21
Oculomotor Disturbance	5.96 ± 2.66	14.62 ± 4.16	22.74 ± 6.1
Disorientation	1.99 ± 1.99	5.97 ± 2.4	19.89 ± 7.11

Each of the four ANOVA conducted for the SSQ analysis reached significance (Table 5.23), and post hoc comparisons are summarised in Table 5.24. Within the subscales of total score and disorientation, it is found that HMD-VR significantly increase cybersickness ratings over baseline and DB-VR presentations. Within subscale combined nausea, it is both DB-VR and HMD-VR that significantly increase reported cybersickness symptoms over baseline. All post hoc comparisons within the subscale of oculomotor disturbance reached significance, finding that both DB-VR and HMD-VR significantly increased cybersickness measures above baseline, and that HMD-VR significantly increased cybersickness above DB-VR. Within Combined Nausea, it was found that both DB-VR and HMD-VR evoked significantly higher ratings of cybersickness than baseline.

Table 5.23: *Main effects and interactions of the 4 ANOVA conducted for the SSQ subscale scores. n=14.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Total Score	9.15	2	26	**<0.001***	0.41
Combined Nausea	8.17	2	26	**<0.001***	0.39
Oculomotor Disturbance	6.8	2	26	**<0.001***	0.34
Disorientation	6.57	1.12	14.55	**0.02***	0.34

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

Table 5.24: *Post hoc comparisons of each SSQ ANOVA conducted between display conditions. n=14.*

Comparison	Cond 1	M	SEM	Cond 2	M	SEM	t	df	M-Diff	Std. Err	P
Total Score	Baseline	5.34	2.24	DB-VR	12.02	3.42	-0.92	13	-3.41	3.69	0.37
	Baseline	5.34	2.24	HMD-VR	24.84	6.21	-3.31	13	-16.35	4.94	**0.01***
	DB-VR	12.02	3.42	HMD-VR	24.84	6.21	-3.18	13	-12.95	4.08	**0.01***
Combined Nausea	Baseline	4.77	2.18	DB-VR	8.18	2.8	-2.51	13	-8.66	3.45	**0.03***
	Baseline	4.77	2.18	HMD-VR	21.12	5.21	-3.67	13	-16.78	4.58	**<0.01***
	DB-VR	8.18	2.8	HMD-VR	21.12	5.21	-1.5	13	-8.12	5.42	0.16
Oculomotor Disturbance	Baseline	5.96	2.66	DB-VR	14.62	4.16	-2.28	13	-3.98	1.74	**0.04***
	Baseline	5.96	2.66	HMD-VR	22.74	6.1	-2.86	13	-17.9	6.26	**0.01***
	DB-VR	14.62	4.16	HMD-VR	22.74	6.1	-2.25	13	-13.92	6.19	**0.04***
Disorientation	Baseline	1.99	1.99	DB-VR	5.97	2.4	-2.03	13	-6.68	3.28	0.06
	Baseline	1.99	1.99	HMD-VR	19.89	7.11	-3.81	13	-19.5	5.11	**<0.01***
	DB-VR	5.97	2.4	HMD-VR	19.89	7.11	-2.45	13	-12.82	5.24	**0.03***

Cond=Condition, M=Mean, SEM=Standard Error of the Mean, t=T-value,
df=Degrees of Freedom, M-Diff=Mean Difference, Std. Err= Standard Error, P=Significance
Bold print and * indicate statistically significant differences, $p<0.05$.

5.3.1.8) Methods of Remembering the Maze

Thirteen of the 14 participants included in the analysis reported using a strategy to remember the maze path. Nine participants reported remembering groups of 4 turns. Two participants reported remembering the whole route, with 1 participant later reporting using the names of characters from a TV show to remember the turns. One participant reported remembering every right turn by the turn number in the sequence (i.e. counting the turns during navigation, and remembering which of those number was a right turn). One participant reported imagining a top-down view of the maze developed during guided navigation.

5.3.2) Electroencephalography Results

The grand average absolute power of theta power in the Fz electrode for the learning, recall and guided conditions for all combinations of display type and maze difficulties can be seen in Figures 5.8, 5.9 and 5.10 respectively. The grand average

alpha absolute power in the Pz electrode for the learning, recall and guided conditions for all combinations of display type and maze difficulties can be seen in Figures 5.11, 5.12 and 5.13 respectively.

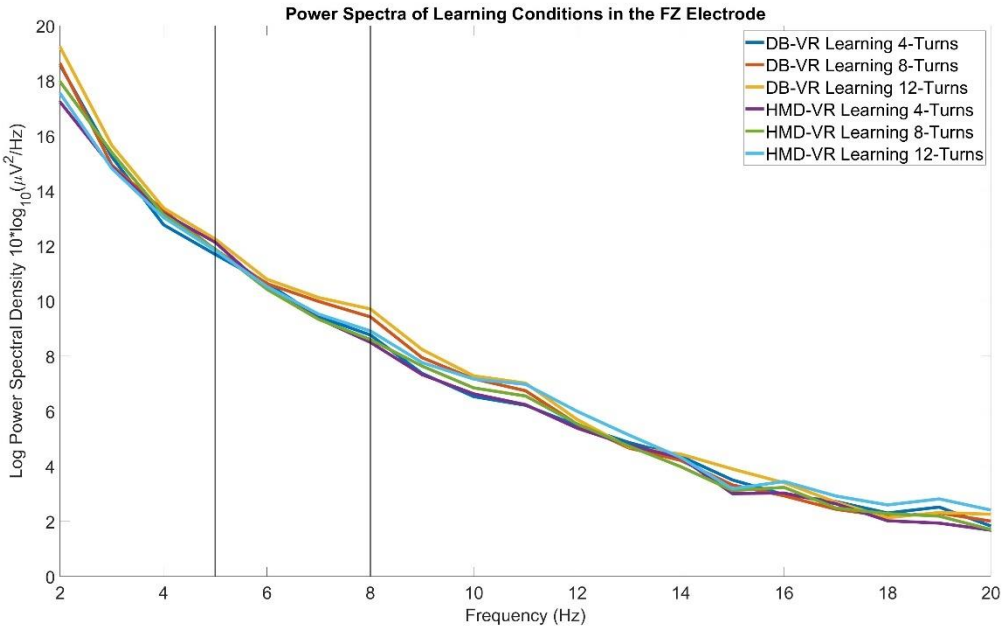


Figure 5.8: *A logarithmic periodogram of the Fz electrode for the learning condition across difficulties and displays. The theta frequency band (5-8Hz) is identified between the horizontal lines.*

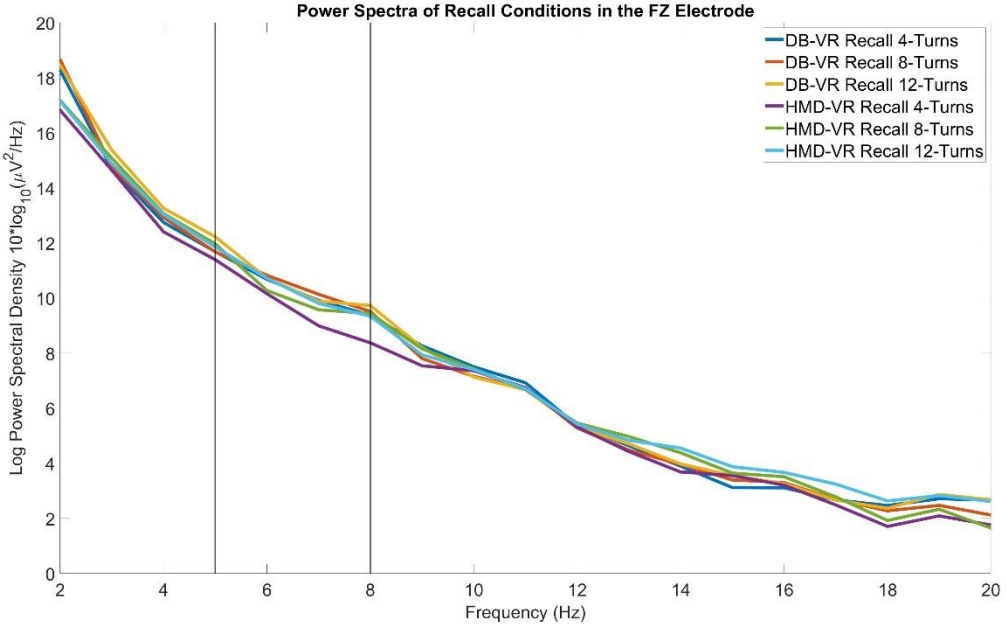


Figure 5.9: *A logarithmic periodogram of the Fz electrode for the recall condition across difficulties and displays. The theta frequency band (5-8Hz) is identified between the horizontal lines.*

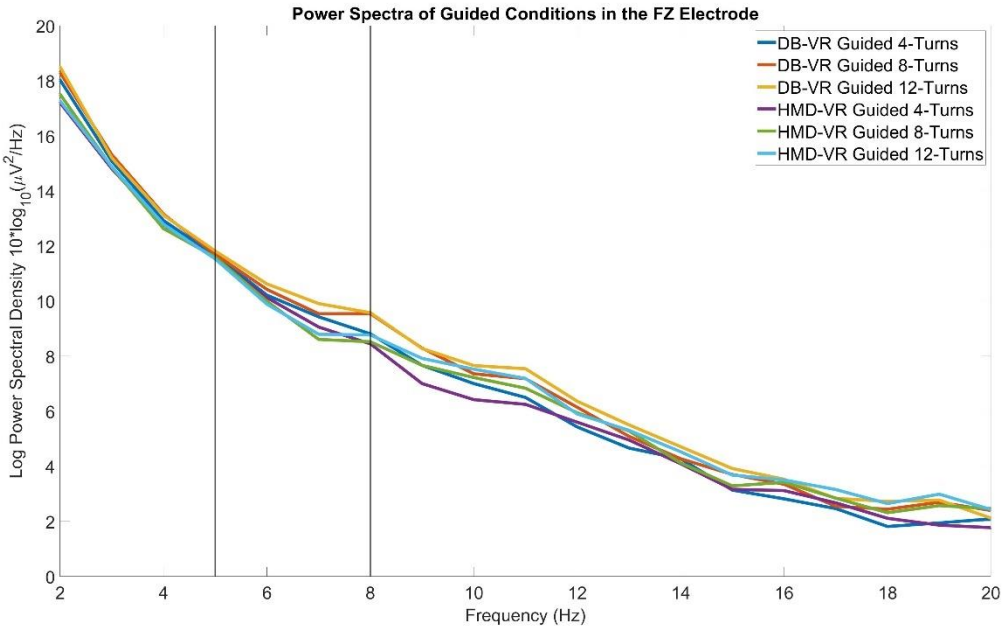


Figure 5.10: *A logarithmic periodogram of the Fz electrode for the guided condition across difficulties and displays. The theta frequency band (5-8Hz) is identified between the horizontal lines.*

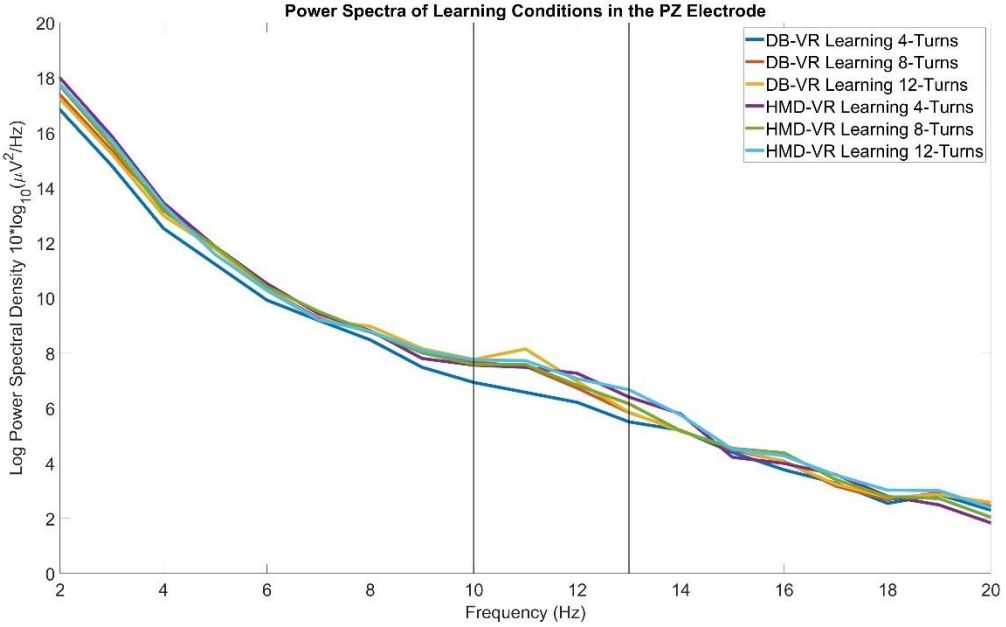


Figure 5.11: *A logarithmic periodogram of the Pz electrode for the learning condition across difficulties and displays. The alpha frequency band (10-13Hz) is identified between the horizontal lines.*

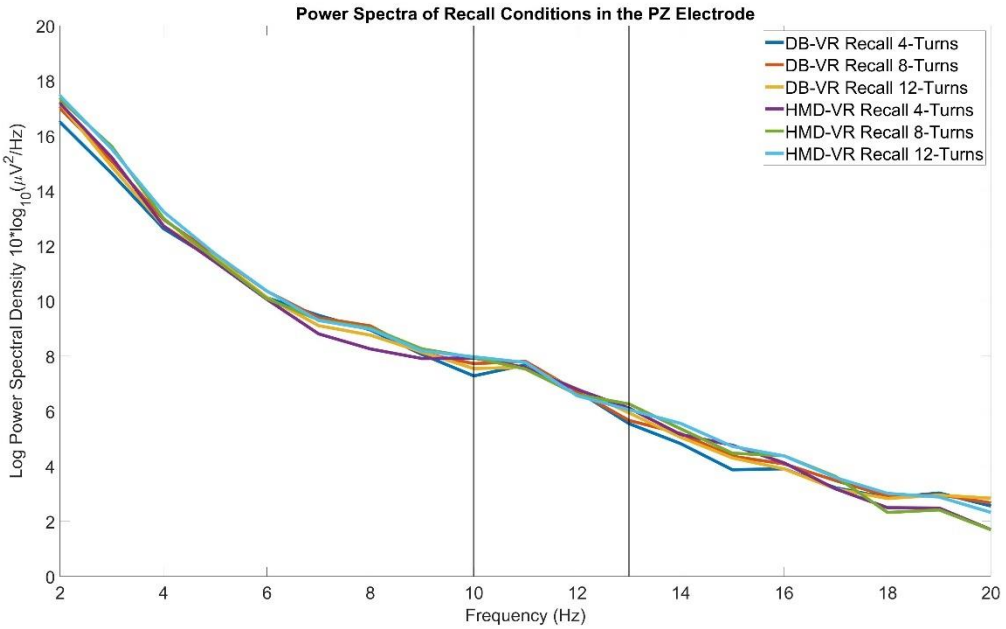


Figure 5.12: *A logarithmic periodogram of the Pz electrode for the recall condition across difficulties and displays. The alpha frequency band (10-13Hz) is identified between the horizontal lines.*

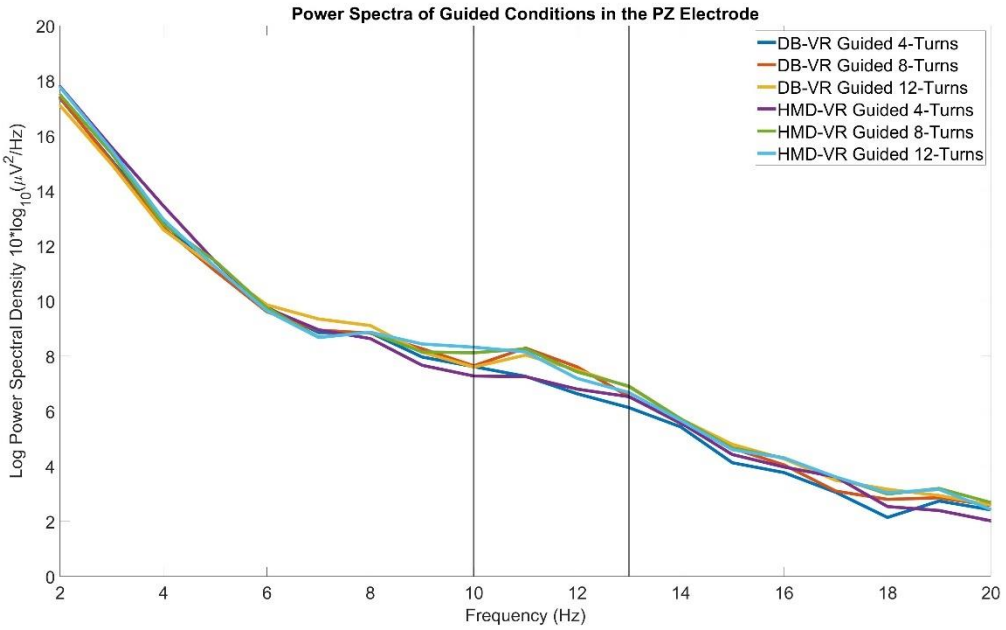


Figure 5.13: *A logarithmic periodogram of the Pz electrode for the guided condition across difficulties and displays. The alpha frequency band (10-13Hz) is identified between the horizontal lines.*

5.3.2.1) Theta Absolute Power

The mean absolute theta power and SEM for each condition included in the analysis is shown in table 5.25. For the 2x3x3 repeated measures ANOVA analyses of theta absolute power (table 5.26) during recall and learning maze navigation, Mauchly's Test of Sphericity is violated for the main effect of display condition x difficulty level ($\chi^2(9)=18.92$, $p=0.03$, $\varepsilon=0.56$), and adjusted with the Greenhouse-Geisser correction. The main effect of condition is statistically significant ($F(2, 26)=4.36$, $p=0.023$, $\eta p2=0.25$), but no post hoc comparison reaches significance.

Table 5.25: *The mean and SEM of the theta band absolute power in the DB-VR and HMD-VR display methods, the learning, recall and guided conditions, and the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

Display	**DB-VR**			**HMD-VR**		
Difficulty	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Learning	10.62 ± 1.23	12.5 ± 1.96	12.63 ± 1.72	10.61 ± 1.44	11.09 ± 1.54	10.97 ± 1.32
Recall	11.12 ± 1.37	12.13 ± 1.54	12.42 ± 1.71	10.79 ± 1.48	11.78 ± 1.73	11.78 ± 1.67
Guided	10.89 ± 1.37	11.69 ± 1.77	11.6 ± 1.5	9.83 ± 1.27	9.91 ± 1.31	9.92 ± 1.34

All values are given as $\mu V^2/Hz$ +/- SEM

Table 5.26: *Main effects and interactions of the theta absolute power ANOVA. n=14.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Display	0.95	1	13	0.349	0.07
Condition	4.36	2	26	**0.023***	0.25
Difficulty level	1.75	2	26	0.194	0.12
Display x Condition	1.19	2	26	0.321	0.08
Display x Difficulty level	0.67	2	26	0.522	0.05
Condition x Difficulty level	0.55	3.17	52	0.659	0.04
Display x Condition x Difficulty level	0.48	2.26	52	0.643	0.04

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

5.3.2.2) Alpha Absolute Power

The mean absolute alpha power and SEM for each condition included in the analysis is shown in Table 5.27. For the 2x3x3 repeated measures ANOVA analyses of alpha absolute power during recall and learning maze navigation, Mauchly's Test of Sphericity was violated for interactions between condition x difficulty level ($\chi^2(9)=50.92$, $p\leq0.001$, $\varepsilon=0.39$) and between display x condition x difficulty level ($\chi^2(9)=23.85$, $p\leq0.001$, $\varepsilon=0.55$). All sphericity violations were adjusted with the Greenhouse-Geisser correction. No main effect or interaction in the alpha band analysis reached significance (Table 5.28).

Table 5.27: *The mean and SEM of the alpha band absolute power in the DB-VR and HMD-VR display methods, the learning, recall and guided conditions, and the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

Display	**DB-VR**			**HMD-VR**		
Difficulty	4-Turn	8-Turn	12-Turn	4-Turn	8-Turn	12-Turn
Learning	4.93 ± 0.62	5.7 ± 0.58	5.97 ± 0.69	5.94 ± 0.93	5.84 ± 0.66	6.08 ± 0.72
Recall	6.25 ± 1.32	6.04 ± 0.93	5.51 ± 0.68	6.06 ± 0.93	6.05 ± 0.91	6.23 ± 0.94
Guided	5.69 ± 0.68	6.42 ± 0.67	6.52 ± 0.81	5.63 ± 0.59	6.89 ± 0.93	6.68 ± 0.83

All values are given as $\mu V^2/Hz$ +/- SEM

Table 5.28: *Main effects and interactions of the alpha absolute power ANOVA. n=14.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Display	1.25	1	13	0.284	0.09
Condition	2.47	2	26	0.105	0.16
Difficulty level	2.81	2	26	0.079	0.18
Display x Condition	0.13	2	26	0.876	0.01
Display x Difficulty level	0.06	2	26	0.938	<0.001
Condition x Difficulty level	0.89	1.57	52	0.402	0.06
Display x Condition x Difficulty level	1.27	2.2	52	0.299	0.09

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

5.3.2.3) Theta/Alpha Ratio

The mean TAR and SEM for each condition included in the analysis is shown in table 5.29, and visualised for the learning condition in Figure 5.14, the recall condition in Figure 5.15, and the guided condition in Figure 5.16. For the 2x3x3 repeated measures ANOVA analyses of the TAR during navigation (Table 5.30), Mauchly's Test of Sphericity was violated for interactions between condition x difficulty level ($\chi^2(9)=20.89$, p=0.01, ε=0.63) and between display x condition x difficulty level ($\chi^2(9)=17.53$, p=0.04, ε=0.6), and is adjusted by the Greenhouse-Geisser correction.

Table 5.29: *The mean and SEM of the theta/alpha ratio in the DB-VR and HMD-VR display methods, the learning, recall and guided conditions, and the 4-turn, 8-turn and 12-turn maze difficulties. n=14.*

Display	**DB-VR**			**HMD-VR**		
Difficulty	**4-Turns**	**8-Turns**	**12-Turns**	**4-Turns**	**8-Turns**	**12-Turns**
Learning	2.34 ± 0.3	2.31 ± 0.33	2.31 ± 0.28	1.93 ± 0.23	1.99 ± 0.24	1.98 ± 0.24
Recall	2.28 ± 0.3	2.39 ± 0.37	2.44 ± 0.35	1.98 ± 0.25	2.09 ± 0.32	2.04 ± 0.25
Guided	2.09 ± 0.3	1.98 ± 0.3	1.99 ± 0.3	1.84 ± 0.22	1.6 ± 0.21	1.63 ± 0.22

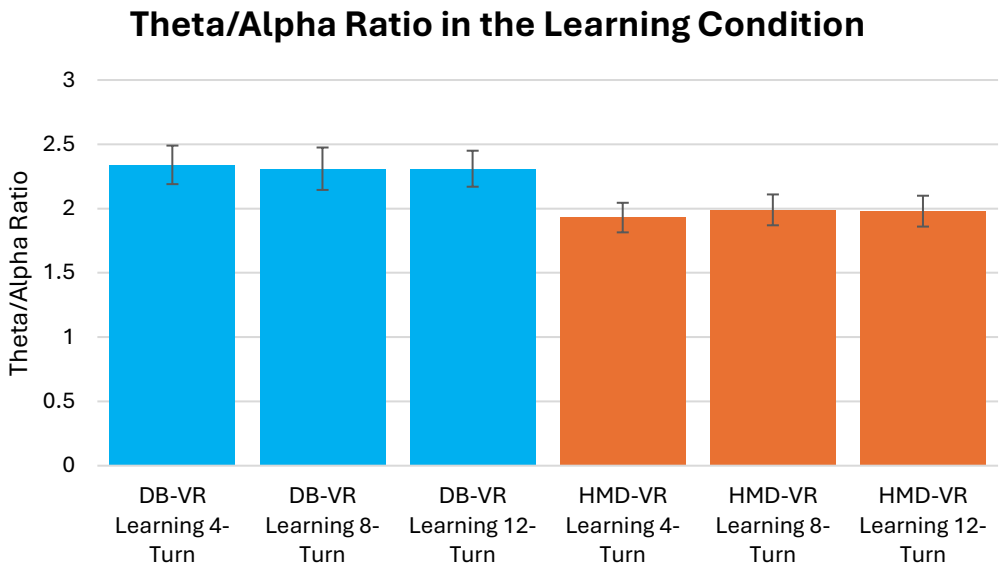


Figure 5.14: *A bar chart of the theta/alpha ratio during the learning condition across both displays and all difficulties. Error bars show the standard error of the mean. n=14.*

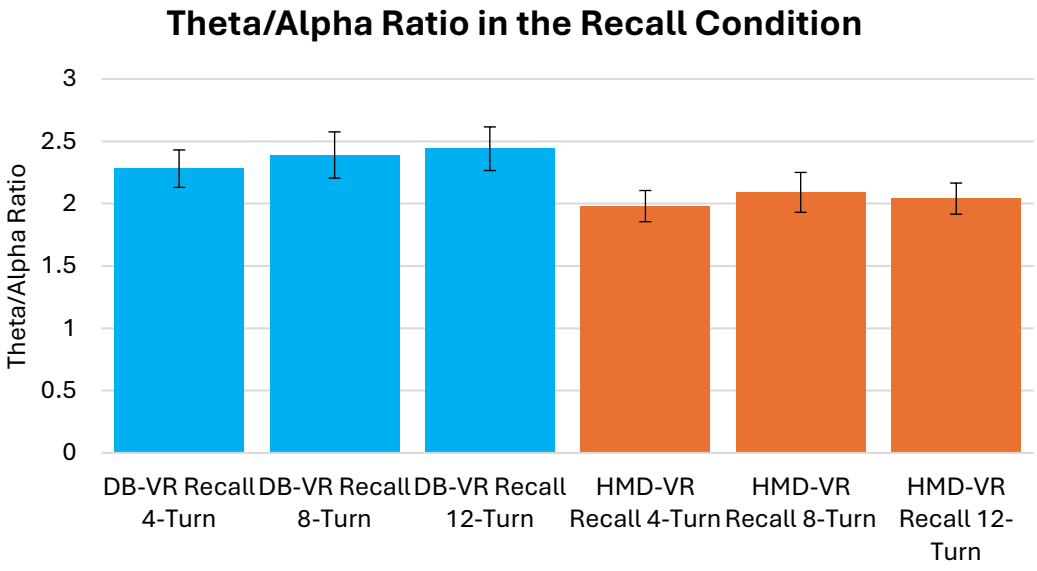


Figure 5.15: *A bar chart of the theta/alpha ratio during the recall condition across both displays and all difficulties. Error bars show the standard error of the mean. n=14.*

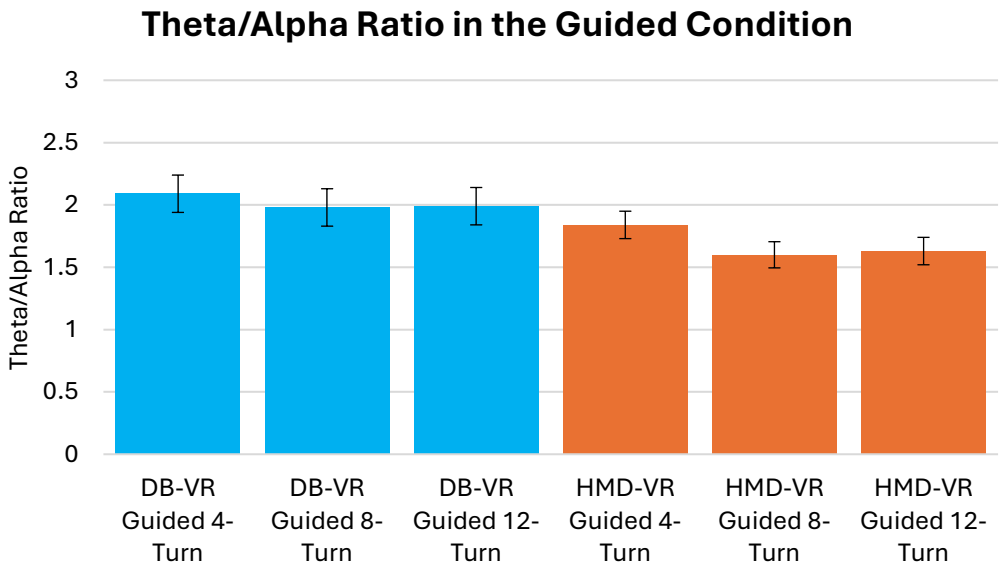


Figure 5.16: *A bar chart of the theta/alpha ratio during the guided condition across both displays and all difficulties. Error bars show the standard error of the mean. n=14.*

The main effect of condition reaches significance (Table 5.30; Figure 5.17), with post hoc tests finding that the guided condition (M=1.86 ± 0.22) resulted in a lower TAR than the learning condition (M=2.14 ± 0.25) (t(13)=0.29, p=0.01, d=<0.001) and the recall condition (M=2.21 ± 0.28) (t(13)=0.35, p=0.01, d=<0.001).

Table 5.30: *Main effects and interactions of the theta/alpha ratio ANOVA. n=14.*

Within Subjects Effect	F	df	df (error)	p	η_p^2
Display	2.97	1	13	0.108	0.19
Condition	12.18	2	26	**<0.001***	0.48
Difficulty level	0.03	2	26	0.975	<0.001
Display x Condition	0.03	2	26	0.975	<0.001
Display x Difficulty level	0.05	2	26	0.955	<0.001
Condition x Difficulty level	1.27	2.53	52	0.299	0.09
Display x Condition x Difficulty level	0.26	2.39	52	0.807	0.02

F=F Value, df=Degrees of Freedom, p= Significance, ηp2=Partial Eta Squared
Bold print and * indicate statistically significant differences, p<0.05.

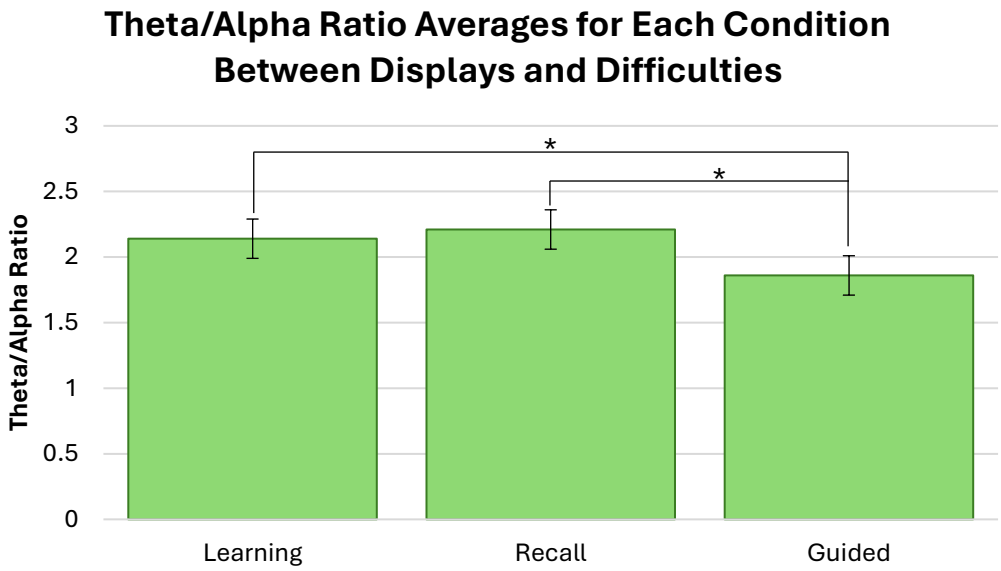


Figure 5.17: *A bar chart of the average theta/alpha ratio for displays and difficulties within each navigation condition. Significant differences are marked with a *. Error bars show the standard error of the mean. n=14.*
**Significant differences between conditions, p<0.05.*

5.4) Discussion

This working memory study examined spatial navigation performance during a route learning and recall maze task, comparing within and between HMD-VR and DB-VR presentation using behavioural and EEG measures. Participants were guided through a T-junction maze consisting of a series of left/right turn decisions with instruction to learn the route, before navigating the same maze unguided. Upon completion of the active navigation trials, the participant was instructed to passively follow the same guided path without attempting to learn the route. The mean absolute power for the frontal-central theta and parietal-central alpha EEG frequency bands, and the TAR between these, were extracted from each successful run of a maze.

The analysis of the behavioural results found that completion time of successful runs, subjective ratings of difficulty, and average number of wrong turns did not differ between displays. It was found that, whilst overall the shorter mazes were completed quicker, the more difficult mazes were completed quicker when restricting completion time to the final 4 turns of the maze (the ‘new’ part of the maze the participant had to

learn in addition to maintaining the previously navigated route). It may be possible that as the 4-turn learning condition was the first conducted in the maze task following the familiarisation period, participants were still familiarising themselves with the input method and thus navigated slower. Subjective ratings of difficulty increased with difficulty levels, with participants rating the 4-turn mazes as easier than the 8-turn, which in turn were rated easier than the 12-turn maze. The SSQ results found that HMD-VR increased symptoms of cybersickness over baseline across all subscales, and over DB-VR in the subscales of total score, oculomotor disturbance, and disorientation.

The EEG results found that the actively navigated learning and recall conditions resulted in a significantly higher TAR than the passively navigated guided condition, indicating active navigation resulted in a higher level of WML [129]. No significant differences between the TARs of displays or difficulties were found. No comparisons within absolute theta band or alpha band activities reached significance.

Despite the interest in utilising HMD-VR in spatial navigation research [373], to this researcher's knowledge this study is the first to successfully utilise the HS-HMD-VR Vive Pro display device in a full maze spatial navigation task. It is demonstrated in this study that HS-HMD-VR can be utilised as a tool to investigate spatial navigation, as participants successfully completed a 12-turn T-junction in the immersive VE.

5.4.1) Maze Tasks and Spatial Navigation High-Specification Head-Mounted Display Virtual Reality

The results in this study find no differences in subjective difficulty ratings or objective task performance differences between HS-HMD-VR and DB-VR presented T-junction mazes, indicating there was no difference in WML between displays. Previous spatial navigation tasks comparing between HMD-VR and alternative displays have also reported no differences between navigation performance [383]. A potential explanation for the behavioural findings is participant age, as Plechatá et al. [354] found that older participants had reduced spatial navigation outcomes in an HMD-VR

supermarket learning and recall task compared to DB-VR. When restricted to ‘young’ participants with a similar mean age to the current study (23.3 to 25.4 years old), there was no difference in spatial navigation outcomes between display types.

An alternative explanation for the lack of difference between conditions is identified in Murcia-López & Steed [146], who compared between spatial navigation task performance in simple low-fidelity and realistic high-fidelity VEs. They found that when using high-fidelity VEs, task performance in HMD-VR outperformed DB-VR. However, when a low-fidelity VE was used, which is visually similar to the type of textures used in the current maze task (Figure 5.18), there was no difference of performance between display conditions. It is therefore possible that a more ‘realistic’ maze presented in HMD-VR would be more beneficial to spatial navigation.

a) Low-fidelity virtual environment

b) High-fidelity virtual environment

c) Real world environment

Figure 5.18: *Image taken from Murcia-López and Steed [146] displaying the 3 types of environments used in their study. The low-fidelity VE (a) has visual similarities to the walls and floors utilised in the current maze task (Figure 5.3), whereas the high-fidelity VE (b) recreation of the real-world environment (c) contains more texture and lighting detail.*

Individual differences between participants may have also been a factor in the WML experienced in HMD-VR navigation. Marraffino et al. [382] found that, on average, there was no difference in spatial navigation outcomes between HMD-VR and DB-VR presentation. However, when distinguishing between low and high video-game experience groups, comparing those who play weekly or more frequently to those who do not, they found that high video-game experience participants performed better across both displays. Moreover, high video-game experience participants

performed better in the HMD-VR condition compared to the DB-VR condition. However, video-game experience has been previously found to have a negative impact on working memory learning outcomes in non-spatial navigation tasks presented in HMD-VR [190]. Whilst the exact benefits of prior video-game experience to working memory processes in HMD-VR is currently uncertain, it could be a factor in the findings reported by this study. In the current maze task, only 5 of the participants included in the final datasets reported playing less than 1 hour of video games per week, with the majority datasets included coming from participants who could be categorised as having high video-game experience. It is possible that experienced VSWM performance was impacted by participant experience (or lack thereof) with adjacent technology, minimising the differences between displays.

The input method used to control virtual movement in the present study was a joystick-based gamepad controller. When the joystick on the gamepad was pushed away from the centre resting point, a participant's avatar in the VE would begin travelling in the direction the joystick is pushed until it is released. Participants were also seated in a fixed chair to discourage head turning during HMD-VR navigation, as the gamepad also controlled rotation in the VE through a second joystick. The same input method was used for both DB-VR and HMD-VR displays in Srivastava et al.'s [166] free-exploration and cognitive map recall task, which reported that HMD-VR presentation increased WML and cybersickness symptoms relative to DB-VR. The researchers argue that, because continuous locomotion does not provide the same physical motion information from walking, the benefits of using HMD-VR is negated or reversed. However, studies utilising button-based navigation have found improved learning outcomes in the HMD-VR condition with only rotational real-world movement [381]. Therefore, full physical motion is not required to benefit from HMD-VR in spatial navigation.

In the current study, HMD-VR usage increased cybersickness compared to baseline and DB-VR usage, which potentially increased experienced WML [165] and negated the benefits of HMD-VR to spatial navigation. Moreover, cybersickness was cited as the reason all 7 (~25%) participants who exited the study early were unable to

complete the experiment. It is possible that that the input method used in the current study contributed to the levels of cybersickness experienced during HMD-VR spatial navigation. Comparisons between input methods in HMD-VR have found that the use of controller-based continuous movement can increase measures of cybersickness compared to methods such as teleporting and physical walking [147,402,403]. It is therefore possible that, if balanced comparisons between displays are not required, symptoms of cybersickness may be reduced if more realistic movement methods are utilised, such as walking in place to emulate the real-world walking [404].

5.4.2) Absolute Power Analysis in High-Specification Head-Mounted Display Virtual Reality

The EEG captured for this experiment isolated the absolute power of the theta and alpha frequency bands in the frontal-central and parietal-central electrodes respectively. The only statistically significant difference in EEG responses found that passive guided navigation resulted in a smaller TAR compared to the active learning and recall conditions. The larger TAR found during active navigation suggests that participants were under a higher WML during these conditions [129], as would be expected whilst conducting a task requiring attention, learning and decision making. However, no other comparisons between display type, navigation condition, or difficulty levels reach statistical significance when using measures of the absolute power of alpha band activity, absolute power in theta band activity, or the TAR. The results therefore suggest that WML did not differ between displays, during any active navigation condition, or between any difficulty level. This is partially supported by the behavioural results, which also found no differences in completion time or number of wrong turns between difficulty levels or displays.

5.4.2.1) Theta and Alpha Band Activity During Navigation

Spatial navigation has long been associated with theta-band activity across a range of paradigms [70,367,394]. In maze tasks, theta activity has been found to increase during higher levels of VSWM load, such as during active navigation, recall and when navigating more complex mazes [70,369,370,372]. Moreover, both alpha and

theta activity have been associated with VSWM tasks presented between levels of immersiveness. It has been shown that theta increases when using a 3D display [371] and alpha decreases when utilising higher immersion single-wall CAVE compared to DB-VR display methods [370], similar to increases in WML [405–407]. However, in the current study, no differences in theta activity were found between displays, difficulties or conditions. Moreover, theta during active navigation was not significantly different than during passive navigation.

A potential explanation for the absence of theta activity differences is relative difference in theta EEG oscillatory activity between participants. Kober & Neuper, [385] report that, whilst theta activity increases for both sexes during spatial navigation, increased theta activity is associated with increased spatial navigation performance in female participants, but decreased spatial navigation performance in male participants. Moreover, it is also known that there are differences of at-rest midline theta and alpha power between sexes of participants, including those similarly aged to those recruited in the current study [408]. Sex could therefore be an important variable to consider when interpreting the results found in the study, as 10 of the 14 datasets included in the analysis were from male participants. It is therefore possible that the weighting towards male participants reduced the overall theta power, however the small sample sizes, particularly when wanting to compare against the female sample, prevents formal analysis.

Alpha band activity, whilst less commonly reported in spatial navigation investigations than theta band activity, has been reported to be sensitive to maze length in spatial navigation [409]. Nguyen Do et al. [376] find that alpha band activity desynchronises both during straight line navigation, and when translating allocentric into egocentric spatial information during navigation. Whereas a large portion of the current maze task involves travelling in straight lines between junctions, no differences in alpha are found between conditions in the present study. Moreover, only a single participant reported using an allocentric method of remembering the maze (reporting they imagined looking down on a representation of the maze), whereas the majority of participants remembered partial or complete turn sequences.

It is therefore unlikely that the alpha activity being comparable between difficulties and between displays results from balanced the use of allocentric and egocentric recall methods between participants.

It is possible that cybersickness experienced by the participants influenced the alpha and theta band activity recorded. There have been several studies which have assessed cybersickness symptoms and EEG frequency bands, but no consensus on how cybersickness effects EEG frequency band activity has been reached. Jang et al. [410] found alpha band activity desynchronises during symptoms of cybersickness, whereas Naqvi et al. [411] and Krokos & Varshney [412] report frontal theta and parietal alpha activity increased with cybersickness symptoms. Whilst it may contribute to the EEG results, cybersickness is unlikely the complete reason for the absolute power findings. If the alpha and theta activity did change with cybersickness, a difference between the display conditions reporting significantly different levels of cybersickness symptoms would be expected, but such a difference is not found. Therefore, whilst it is important to consider cybersickness and the effect on WML and EEG recording, it cannot explain the current EEG findings alone.

5.4.2.2) The Theta/Alpha Ratio and the Working Memory Load Perspective

Participants in the present study were tasked with remembering increasing number of turn sequences, designed to evoke different levels of WML. This increasing turn sequence is analogous to digit span tasks, where participants must remember increasingly-sized lists of numbers [405,413]. In WML tasks, frontal theta and parietal alpha activity are well documented to increase and decrease respectively under heightened levels of load [121,134,405]. Therefore, the TAR, which is sometimes called the ‘cognitive load index’, was selected as a measure of WML [338]. The sensitivity of the TAR to WML is demonstrated by finding the significant increase in TAR during active learning and recall navigation compared to the passive guided. Moreover, the TAR has been used to investigate factors related to experienced immersion in VR research [129], and thus should be suitable for identifying differences in experienced WML between display conditions and difficulties.

The lack of significant differences between display types is not considered a negative result for HMD-VR usage. As with the Chapter 2 systematic review, the objective of this study is not to show that HMD-VR is ‘better’ than DB-VR for working memory tasks such as spatial navigation, but to demonstrate how it is an appropriate tool for neuropsychological research. In the current study, there is no difference between displays on any measure, suggesting that the use of HMD-VR did increase WML in comparison to DB-VR.

The most likely explanation for the absence of differences in EEG and behavioural measures between conditions and difficulty levels is that the 4-turn, 8-turn and 12-turn maze lengths did not differ enough in difficulty to induce distinct levels of WML. The decision to utilise 4, 8 and 12 turns was based on the working memory capacity of 7+/-2 [90]. A length of 4 turns was selected to probe under the working memory capacity range, 8 turns was selected to probe within the working memory capacity range, and 12 turns was selected to exceed the working memory capacity range by being over the 9 item maximum limit. There is precedence for expecting a difference between the 4 and 12 turn conditions, as previous comparisons between 6 and 12 turns have resulted in measurable differences in theta EEG [369].

The behavioural results found no significant differences in the number of wrong turns, indicating the ‘hardest’ mazes did not differ in performance from the easiest despite the higher number of turns. Participants overcame their working memory capacity by chunking turns together into groups of 2 or 4, reducing the WML necessary between runs (though 2 participants reported remembering the entire path in sequence). Moreover, the average subjective difficulty rating of the 12-turn course (on a scale of 1 = very easy to 10 = very hard) was 5, and only one participant was excluded for having no successful runs in a 12-turn recall trial. These findings indicate that the majority of participants experienced relatively little issues completing the course, and were not cognitively overloaded nor particularly challenged during navigation. No participant reported using the corridor length as a guide or aid to remembering turns, so it is believed the different length corridors were unnoticed or ignored.

The present study does not find any difference in EEG theta frequency band activity between learning and recall during spatial navigation. It has previously been reported that recall during spatial navigation results in higher levels of theta activity compared to learning during spatial navigation [70,369]. More recently, a comparison between learning and recall during spatial navigation found no difference in theta band activity [370], aligning with the findings of the current study.

A potential explanation for the lack of EEG differences between learning and recall trials is the amount of information held online in working memory during each form of navigation. After the first of the 3 learning runs of the maze completed for each for learning condition, participants should be holding some form of the complete sequence in their working memory, totalling the complete sequence by the end of the learning blocks. It is therefore possible that only the first of each maze learning run was actually 'learning', and subsequent runs was maintenance and reinforcement of the information in working memory. We know participants held this information as they were able to subsequently complete the recall trial, which required the same amount of information being maintained in working memory. It may therefore be worthwhile restricting data extraction to the first learning condition maze run only, or comparing the first and final maze run to identify differences in TAR.

5.4.3) Limitations and Future Directions

The number of participants included in the final analysis for the present study was within the standard range used for EEG research [183], with 14 participants previously being used in similar EEG-based maze paradigms [70]. However, other spatial navigation experiments using frequency analysis have included ~30 datasets in the final analysis [370], including those utilising HMD-VR [394]. The lower number of participants may therefore have reduced the power of any statistical analysis performed, potentially limiting the power of the EEG findings reported or resulting in a type 2 error. Future studies should therefore ensure more participants are recruited, and minimise the number of participants who exit the study by reducing factors which can result in cybersickness symptoms. Future studies examining methods of reducing cybersickness in HMD-VR, which was the principle cause for

the loss of datasets in the current experiment, must conducted to better understand how to prevent data loss.

Due to the number of participants exiting the study early or otherwise having data excluded, Maze-A was used in the HMD-VR condition 10 times and the DB-VR condition 4 times in the included datasets. Care was taken to implement rules when designing mazes to prevent one maze being 'harder' than the other, however comparison of the behavioural results suggest that Maze-B was more difficult than Maze-A. Participants completed Maze-B significantly slower than Maze-A, and subjectively rated the 12-turn difficulty level of Maze-B as significantly more difficult than 12-turn difficulty level of Maze-A. However, not every behavioural measure finds the mazes differed in difficulty. There was no difference in the average number of wrong turns between mazes, and the 4-turn and 8-turn version of each maze did not differ between mazeIDs. It is therefore uncertain to what degree the mazes contributed to the results found, as the 'easier' maze was used in the HMD-VR condition more frequently, though not enough to result in significant differences between displays.

It is advised that future studies take care when designing maze paradigms and implement stricter requirements on maze navigation paths to prevent differences in maze difficulties influencing the results found. It is also recommended that researchers conduct a behavioural pilot study to compare the difficulties of potential mazes for use in spatial navigation research.

5.5) Conclusions

This study has demonstrated that HS-HMD-VR can be successfully utilised in spatial navigation-based learning and recall paradigms using T-junction mazes. Task performance and EEG measures were compared between HS-HMD-VR and DB-VR mazes, utilising balanced VEs and input methods. Absolute power recordings of theta and alpha frequency band activity from the frontal-medial and parietal-medial electrodes respectively were acquired using EEG in both display conditions. Whilst it

was found that active navigation, such as learning and recall, resulted in a larger TAR than passively guided navigation, no other differences between display conditions were found within the behavioural and EEG results. Moreover, no behavioural differences between task performance were found between maze difficulties. These results indicate that HS-HMD-VR does not increase WML over DB-VR during spatial navigation.

Reported levels of cybersickness symptoms were higher in post-HMD-VR usage compared to baseline and post-DB-VR usage. Whilst several participants withdrew from the experiment early due to cybersickness, the majority of participants successfully completed the experiment, showing how HS-HMD-VR is suitable for use in the study of spatial navigation. Moreover, those who withdrew primarily consisted of participants with no prior experience with HMD-VR, indicating that the use of certain HS-HMD-VR VEs should only be utilised with individuals who are familiar with the display method.

Whilst maze tasks have been previously utilised using HMD-VR [397,398,414] to the researcher's knowledge this is the first HMD-VR and HS-HMD-VR study to use a T-junction maze. Moreover, it is believed to be the first to combine the Vive Pro HS-HMD-VR device and EEG during a learning and recall maze task. It has been demonstrated that participants successfully learned increasing lengths of turn sequences through a T-junction maze in both display conditions using conventional gamepad controls.

Chapter 6) General Discussion, Limitations and Future Directions

6.1) General Discussion

The overarching aim of this thesis was to investigate the combined use of HMD-VR and EEG as a methodology in the study of working memory. HMD-VR devices became accessible and affordable to the general public with the release of the first 'modern' HMD-VR device in 2013, resulting in a renewed interest in the use of immersive display technology in psychological and neuroscience research [5,20,29]. Subsequently, a range of relatively low-immersion smartphone HMD-VR and relatively mid-immersion DB-HMD-VR configurations have been successfully combined with EEG in studies of working memory processes, as reported in the systematic review (Chapter 2). It was also found that high-immersion HS-HMD-VR devices such as the HTC Vive Pro had not been used in comparative working memory research, nor generally in studies utilising EEG ERP measures.

The first major finding of this thesis is that HS-HMD-VR, specifically the HTC Vive Pro, can be used in combination with EEG methodologies. Two working memory studies, an arithmetic task (Chapter 3) and a spatial navigation task (Chapter 5), were successfully conducted using HS-HMD-VR in combination with EEG. The studies used a range of EEG analysis methods to successfully distinguish between levels of WML between conditions within each task. In the arithmetic task, N170 and P300 ERP responses acquired during visual presentation of addition arithmetic questions could be used to distinguish between question difficulties. In the spatial navigation task, a ratio between the absolute powers of the theta and alpha frequency bands, a measure associated with WML [129,338], was found to increase during active learning and recall navigation compared to passive guided navigation. It is also demonstrated in the preprocessing comparison study (Chapter 4) that standard EEG data preprocessing methods can be used to remove line noise, eye-based artifacts and slow drift artifacts associated with the use of HS-HMD-VR.

The second major finding in this thesis is that HMD-VR does not inherently evoke a larger WML compared to DB-VR presentations of a working memory task. The comparison of HS-HMD-VR and DB-VR in the spatial navigation task, which was designed to have minimal differences between the task presentation outside of display method used, found no difference in behavioural or EEG measures of WML between displays on any comparison. The spatial navigation task results build upon the outcomes of the systematic review, which concluded that various non-high specification HMD-VR configurations including smartphone HMD-VR, non-modern DB-HMD-VR and modern DB-HMD-VR also do not inherently increase experienced WML compared to alternative displays during working memory tasks [66,167,173,180]. When WML was reported to increase in the HMD-VR condition in the papers included in the systematic review, the increased load could be attributed to an aspect of the VR configuration separate from the display device itself, such as an inappropriate input method [16] or the low refresh rate of the displayed image [25].

The compatibility of HMD-VR and EEG coupled with the finding that HMD-VR does not inherently increase WML over other display methods opens a range of potential studies which can benefit from the advantages offered by the display method. In research, experiments can utilise HMD-VR to present ecologically valid VEs whilst blocking out external distractions. For example, HS-HMD-VR can be used to place participants into the driving seat of a virtual car within a high-fidelity VE to study dangerous driving conditions in a safe environment. Attention, WML and stress can be compared using EEG and behavioural measures during various distractions, presence of driving assistance tools (automatic gearboxes, acceleration/breaking guideline, etc.), or various weather conditions [24,188,415,416]. The popularity of HMD-VR in recent years has also resulted in an increase in compatible software that have be repurposed for research, for example the logic-based problem solving game ‘Keep Talking and Nobody Explodes’ [417] and the block puzzle game ‘Cubism’ [418]. For the hypothetical driving experiment, the commercially available video game ‘Assetto Corsa’, a racing simulator which provides control over the vehicles, track and weather, has already been used in research comparing HMD-VR and DB-VR presentations [419].

It was also found in this thesis that the combination of HMD-VR and EEG in neuroscience research did not require extensive technical expertise to utilise the methods together. No specialised modifications of either the Vive Pro HMD-VR device or the SPESMedica EEG Sleepcap were required in either the Chapter 3 arithmetic task or the Chapter 5 spatial navigation, as the soft foam-padded electrodes were able to be placed under the HMD. In both experiments conducted in this thesis, commercially available software was used to present the paradigms. Whilst the Garry's Mod software did require modification for the maze study, dedicated software for spatial navigation paradigms [420] or toolkits for experimental data collection in the Unity game engine [421], both of which natively support HMD-VR, are available. With the range of software available now, it is expected that future studies will often have pre-packaged or easy-to-construct solutions available for whatever paradigm that will be employed. Once the data was collected, no specialised knowledge or method was required (outside of understanding general EEG procedures and analysis techniques) to preprocess the EEG data for statistical analysis. The parameters compared for each preprocessing step in the HMD-VR EEG artifact removal study were selected as they have been used both within and outside of HMD-VR research.

The finding that HMD-VR does not increase total WML compared to other display methods also has implications for the practical applications of the display method. As the use of the HMD-VR does not inherently increase WML, VEs tailored to maximise attention and reduce WML can be designed and utilised in a range of scenarios. HMD-VR can facilitate teaching and workplace use by presenting learning materials and virtual office spaces designed in a way to minimise extraneous load, maximising the working memory resources available for the current task. The practical applications of HMD-VR are of particular interest following the increase in remote learning following the COVID-19 pandemic, where being able to present immersive classroom environments could reduce the external distractions reported during online classes [106,422].

6.2) Limitations and Future Directions

6.2.1) Comparisons of Working Memory Load Between Different Head-Mounted Display Virtual Reality Devices

A criticism of the existing literature identified in the systematic review is that the immersiveness of HMD-VR configurations used in research is often overlooked during the interpretation of working memory results. Not considering the effect of immersiveness is particularly problematic when using configurations with relatively low immersiveness, of which initial comparisons have found can increase WML relative to higher immersion configurations [6]. However, the exact relationship between HMD-VR immersiveness and WML is uncertain, as there has been limited comparisons between HMD-VR devices, or between otherwise roughly equivalent devices [423]. A similar limitation applies to the current thesis, in that the conclusions drawn may only apply to the Vive Pro HS-HMD-VR. The Vive Pro used provided a higher immersion over the development-grade DB-HMD-VR and smartphone-HMD-VR devices found in the systematic review. However, as no direct comparisons of WML was conducted between HMD-VR devices within this thesis, nor have the same studies been completed with other HMD-VR configurations, it is difficult to conclude if HS-HMD-VR had any effect on experienced WML.

It is recommended here that future studies examining the effect of HMD-VR on working memory should include comparisons between different HMD-VR configurations. By comparing WML experienced from the same task presented between HMD-VR configurations, the impact that individual factors of each display and the overall immersiveness has on working memory processes can be identified. Additional measures of presence and cybersickness should also be captured during these comparisons to further understand the subjective and negative effects of each display configuration.

An additional benefit of comparison between HMD-VR devices is that differences in the electrical artifacts introduced to the EEG recording can be identified. The comparison between the Vive Pro and Oculus Rift CV1 by Weber et al. [287] has

demonstrated that different devices can introduce different frequency band artifacts, likely resulting from the internal components used in each HMD. It is therefore possible that components in modern standalone HMD-VR devices, such as internal batteries or wireless Wi-Fi and Bluetooth capabilities absent in most DB-HMD-VR devices, may introduce different frequency noise or other artifacts to an EEG recording. Understanding the effect of different HMD-VR configurations on the EEG recording will allow researchers to make informed decisions when selecting which HMD to use in research, and how to preprocess the EEG data to remove the artifacts prior to analysis.

6.2.2) Cybersickness in Head Mounted Display Virtual Reality

The overarching limitation of HMD-VR identified in this thesis is the increased cybersickness whilst using immersive HMD-VR. Participants in both the arithmetic and spatial navigation tasks reported cybersickness symptoms following completion of the respective HMD-VR conditions. Whilst cybersickness is not an exclusive issue for HMD-VR, as it is also found to increase in DB-VR in the Chapter 5 spatial navigation task, the increase in cybersickness symptoms reported was significantly larger following HMD-VR presentation.

Cybersickness can impact both the recording and interpretation of working memory results in studies conducted using HMD-VR and EEG. Cybersickness has been linked to increased WML in HMD-VR conditions [32,141], introducing extraneous load and serving as a confounding factor when interpreting working memory results. Cybersickness has further been found to affect the same P300 ERP component [424] and theta & alpha EEG frequencies bands [410–412] used as measures of WML in this thesis. This overlap between cybersickness and WML EEG measures could make it difficult to distinguish between working memory-related cognitive processes and cybersickness-related neural responses in the captured data. For instance, it is possible that WML would be lower in the HMD-VR condition of the Chapter 5 spatial navigation task if not for the increased cybersickness found in the HMD-VR condition. However, without a way of separating the effect of cybersickness, or through a comparison with a group who did not experience any

cybersickness during spatial navigation, it is impossible to determine if cybersickness had a large enough effect to change the results found.

At its most severe, cybersickness prevented the acquisition of data altogether, as demonstrated by the 1 (out of 22) participant in the Chapter 3 arithmetic task and the 7 (out of 27) participants in the Chapter 5 spatial navigation task who exited the respective studies early due to cybersickness symptoms. The difference between the number of self-excluded participants is likely a result of the task used, as cybersickness has been associated with the 'sensory mismatch' between seeing visual information indicating the body is moving in space when a person is stationary [157–159]. In the spatial navigation task, participants visually experienced continuous locomotion including rotation whilst navigating the VE, but whilst sat looking forward in a chair in the real world. Compared against the arithmetic task where participants were presented a low-visual stimulation VE whilst seated, the increase is unsurprising. However, as one participant exited the arithmetic study early due to nausea, the sensory disconnect cannot be the only explanation for increased cybersickness symptoms.

To prevent data loss and adding extraneous load to working memory, it is important that future HMD-VR research takes steps to minimise cybersickness wherever possible in HMD-VR research. The most basic step that every experiment must take is to ensure the HMD-VR is properly calibrated for each participant, as recommended in Chapter 1 and used in the arithmetic and spatial navigation tasks. In the studies conducted within this thesis, blocks of text that had to be read aloud were used to check that interpupillary distance and HMD position were correct. The blocks of text displayed emphasised that if it was not clear and easy to read, the participant must tell the researcher so it can be corrected. Calibration must include the peripheral areas of vision, as visual distortions in the peripheral vision can exacerbate cybersickness symptoms [186]. By extension, it is recommended that software which fully supports HMD-VR, or has been optimised in a way to reduce cybersickness symptoms [197], is utilised in to prevent visual compatibility issues that may give rise to visual artifacts in the VE.

It is also important ensure the technical specifications of the equipment used can support the smooth presentation of a VE within HMD-VR. Factors including low refresh rate of a screen [187] and high latency, the delay between an action such as head movement being performed in real life and it being reflected in the VE [425], can increase the sensory mismatch and lead to feelings of discomfort. Researchers must ensure prior to the onset of the experiment that no sudden drops in refresh rate, or noticeable delays between actions being performed and the actions being reflected in the VE, are present in the experimental procedure used.

Additional recommendations will be largely paradigm specific. In tasks requiring navigation, joystick-based continuous locomotion, as used in the spatial navigation task, has been found to increase cybersickness symptoms in participants over teleportation-based navigation [403,426–428]. Ideally, the sensory disconnect between visual and vestibular information would be minimised by using walking-based navigation in HMD-VR [429]. Naturalistic methods of navigation that have been previously used in HMD-VR include physically walking in a large real-world space [430], and through peripheral technology such as omnidirectional treadmills which emulates naturalistic locomotion in HMD-VR VEs whilst staying in one location in the real world [431].

Researchers must also be aware of what individual factors between participants may affect cybersickness severity. Recently published research has identified motion sickness susceptibility and video game experience as key predictors of cybersickness [432]. A lack of prior experience with HMD-VR is also found to be associated with cybersickness in the current thesis, dependent on the ‘intensity’ of the VE. In the arithmetic task, which used a ‘gentle’ stationary VE with minimal visual stimuli, only one of the fourteen participants reporting no prior VR experience exited the study. Conversely, all participants in the Chapter 5 spatial navigation task with no prior HMD-VR experience exited the study early due to cybersickness, likely due to the relatively high visual demands of the continuous locomotion used to navigate the mazes.

6.2.3) Scales of Immersiveness to Rate HMD-VR Devices

The advancements in the technical specifications of HMD-VR devices in recent years warrants a review of how HMD devices are categorised. In Chapter 1, four subcategories of HMD-VR were identified of relative immersiveness: HS-HMD-VR, which offers the highest immersion; high-immersion standard consumer-grade DB-HMD-VR; mid-immersion standalone HMD-VR; and low-immersion smartphone HMD-VR. Whilst these sub-categories were appropriate for when the research was conducted, the technical specifications of recent HMD-VR devices overlap between the sub-categories. For example, a standalone Meta Quest 3 HMD-VR device released in 2023 has a 2064x2208 pixel-per-eye resolution, a 110 degree FOV, and a 120Hz refresh rate display. Compared to the at-the-time 'high specification' Vive Pro used in this thesis, which was released in 2018 and has 1440x1600 pixels-per eye, a 110 degree FOV, and an up to 90Hz refresh rate display, the standalone Meta Quest 3 offers a higher level of immersion. The division between a standalone HMD-VR device and a DB-HMD-VR device is also now less clear. Apicella et al. [433] state the standalone HMD-VR used in their experiment was "low end" due to the quality of the VE being restricted by the internal components processing the VE, which cannot currently create the quality of VEs used in DB-HMD-VR. However, recent standalone HMD-VR devices such as the Meta Quest series can also operate as DB-HMD-VR devices through wired or wireless connection to a host computer processing a VE. Wireless connection also has the additional benefit of not requiring physical tethering, removing a physical restriction common for standard DB-HMD-VR use.

Considering the ongoing advancement of HMD-VR technology, it is recommended the way immersiveness is categorised or measured is reviewed to facilitate future comparison between HMD-VR devices in the study of WML. Instead of continuing to use distinct categories of devices, the development of an 'immersiveness scale' that accounts for each aspect of an HMD-VR display relating to immersiveness, including resolution, field of view, degrees-of-freedom, weight, screen latency, and any movement restrictions (i.e. tethering to a computer) is recommended. These factors should in turn be weighted based on how much they effect experienced presence, cybersickness and experienced WML, which can be examined by manipulating each factor within the same headset. Using the device settings, settings within the VR

management software or restrictions coded into a VE, certain parameters such as rendering resolution or degrees of freedom can be restricted, allowing comparisons between specific factors. Each item on the immersiveness scale should also have an upper limit based on human sensorimotor perception where further technical improvements will not be noticeable. For example, a 360-degree field of view will not be of any additional benefit to sensorimotor perception compared to a 270-degree field of view if normal naturalistic looking is only ~200 degrees [434].

Whilst it is important to identify the immersiveness of an HMD-VR configuration, any benefits to WML and other cognitive functions can be reduced or reversed if the immersive experience presented through the display is substandard. Therefore, a VE immersiveness scale examining factors that can affect visuospatial processing within the VE should be developed. The VE immersiveness scale should include factors not wholly dependent on the HMD-VR device, such as frames-per-second of the VE, the latency between input and representation in the VE, and the fidelity of the VE, each of which can be affected by graphical processing power. Peripheral devices such as the appropriateness of input methods of the VE and if and how navigation within the VE was performed should also be considered in the VE immersiveness scale. Some factors included in the scale may be more appropriately measured in broad categories opposed to specific numbers. For example, for the input method in a driving task, a steering wheel would be one-to-one with the real world, physically moving motion controllers to emulate the movements of a steering wheel would be semi-realistic, using a gamepad controller may be unrealistic but suitable (as fine control is offered to acceleration and steering), and a keyboard would be unrealistic and unsuitable (rate of acceleration or turning cannot be finely adjusted). Alternatively, if a scale is not used, the listed factors should still be reported as they pertain to WML and cybersickness and may aid interpretation of the results.

The benefits of the immersiveness scales are threefold. Firstly, it will make researchers more aware of the importance of selecting appropriate HMD-VR configurations for each research paradigm. In particular, the immersiveness scale will help new researchers better understand the technology available and the

differences between configurations. Secondly, examining the differences between HMD-VR configurations will allow a better understanding on how the technology impacts WML, presence and cybersickness, building upon the comparisons suggested in Section 6.2.2. Thirdly, reporting the HMD-VR and VE immersiveness scales will increase the replicability of an experiment by allowing exact specifications of VR configuration and VE to be recreated. If the exact parameters are not available, differences between conditions can still be quantified.

6.2.4) Systematic Review Future Directions

6.2.4.1) Updated Systematic Review

The systematic review was an important early step into broadly understanding both how HMD-VR had been combined with neurophysiological recording methods, and how HMD-VR influences working memory load relative to alternative display methods. The majority of papers included in the final analysis comprised of SB-VR display methods, with some comparisons between real life conditions and none between CAVE or augmented reality display methods. The review collated papers covering a wide range of working memory tasks comparing between displays, and did not focus on any specific working memory related cognition. The inclusive criteria allowed for a broad identification of if and how HMD-VR effects processing compared to other displays.

Technology and research have progressed since the systematic review was conducted, and newly published papers comparing WML in HMD-VR to alternative display methods may provide additional evidence in support or refuting the conclusions reached by the current systematic review. It is therefore recommended that a follow-up review is conducted to explore how the technology and research has progressed over time, and what effect this has had on our understanding of the use of HMD-VR in working memory processes. This review should also capture papers using newer high-immersion HMD-VR devices, potentially allowing investigation of how different levels of immersion within HMD-VR can effect working memory processes.

A follow up review may also not require neurophysiological methodologies as a necessary inclusion criterion, as the behavioural results in the included papers in the systematic review and the objective behavioural measures in the Chapter 5 spatial navigation task largely corresponded with the neurophysiological results.

6.2.4.2) Nomenclature of Head-Mounted Display Virtual Reality

It was found in the systematic review that the nomenclature used for HMD-VR in published research papers is inconsistent, using ‘HMD-VR’ [173,177] ‘immersive virtual reality’ [16,178], ‘VR Headsets’ [176], and simply ‘VR’ [180] interchangeably. Inconsistent or uninformative terminology can make it difficult to identify which HMD-VR configuration was used in an experiment, or if an HMD was used at all.

The argument for standardising VR terminology in published research papers has been made before [435], however a consistent method has yet to be used in the wider academic literature. It is suggested here that a consistent naming structure that can include all forms of VR is used when reporting or discussing VR methodologies. When using a form of VR, the abbreviated display method followed by ‘VR’ should be used to inform readers that a VE was used, and how it was presented. For example, ‘HMD-VR’ would refer to head-mounted display-based VEs, and ‘DB-VR’ for desktop-computer-based VEs. Additional modifiers can be used at the start of the terms to further clarify the methods employed, for example ‘smartphone HMD-VR’. When possible, any relevant details about the VR method should also be reported, including models of display, relevant details about supporting hardware (e.g. the graphics card on a computer generating the VE), and any modifications made to the device.

6.2.5) Arithmetic Task Limitations and Future Directions

Within the arithmetic task, two limitations are identified that should be considered when designing future arithmetic-based studies. The first limitation is that the ERP responses were taken from the onset of the question presentation, which prevented

comparison between presentation modalities and limited the power of the auditory conclusions. Participants immediately had access to all the information required to perform the calculation in the visual condition, as evidenced by the presence of the positive SWC [131,248]. Conversely, the auditory trials took ~1s for the easy condition and ~2s for the hard condition to fully deliver the same information, meaning the question was still being encoded at the time of the ERP recording. As there are most likely different cognitive processes occurring at the time of the ERP recording between modalities, comparisons of ERP components between visual and auditory conditions were inappropriate for this design and were not conducted.

One solution for future research would be to change the paradigm, for example presenting the equation in sequence and taking an ERP recording from the presentation of the operation or second operand [204,221,256]. This has three main benefits, in that it can be used to induce a level of load within the participant through the size of the first operand, can better balance between visual and auditory presentations, and allows more direct comparison with the existing literature. However, it has been previously reported that even within sequential presentation, the stimuli presentation between visual and auditory digits is different due to the time differences required to present the same information, which can prevent balanced comparison [248]. An alternative solution would be to measure ERP responses from the presentation of a potential solution using a verification task [241]. Verification paradigms allow for an additional level of control on WML through how correct the solution is [223,240,245–247]. The solution could also be presented in the same modality (i.e. visually) regardless of how the equation was presented, allowing a more balanced comparison [238,239]. It is further possible to combine these paradigms and take measurements at both the second operand and the solution presentation, leading to increased amount of extractable information.

The second limitation in the auditory experiment is that, due to the audio file generation method used, there is a ~50ms silent period at the start of each auditory question. However, the time windows used for the auditory condition were ~100ms later relative to the visual condition, suggesting that the perceivable onset of the

auditory stimuli was even longer. It is for this reason it is difficult to identify exactly which auditory ERP components have been found in the current study, further preventing comparison between conditions. Future studies should avoid this by more rigorously inspecting each generated audio file to identify where the sound begins in the spectrographic representations of the soundwaves, and when audio can actually be heard.

6.2.6) Preprocessing Comparison Future Directions

Future research should expand upon the EEG preprocessing comparison study by including additional variations of the preprocessing steps examined for data collected during HS-HMD-VR use. Within highpass filtering, additional half-amplitude cut-off frequencies between 0.1Hz and 0.5Hz [310] can be examined to find the optimal trade-off between noise and amplitude reductions. Other methods of removing eye-based artifacts can also be compared, such as the ‘cleanline’ plugin for EEGLab [311] and or similar algorithms that identify and remove ocular artifacts from the continuous waveform.

Additional EEG configurations and preprocessing steps could also be compared in the context of EEG data collected in a HS-HMD-VR study. One example would be to compare between the types of filters used. Highpass and lowpass filtering is performed in this thesis using Butterworth filters, however it has been suggested that FIR filters are more successful at removing 50Hz line noise from the EEG signal [57]. Another example is the reference used during data collection and pre-processing. Each study in this thesis collected data using a physically linked ears reference joined by a wire, which should be compared against other referencing methods that have been used in combined HMD-VR and EEG experiments. Methods including single-ear referencing [176,177], linked-mastoid referencing [16,63], and the offline average re-reference [175] have previously been reported in the HMD-VR literature, though how these effect data collected using HMD-VR is uncertain. Comparison with single ear electrode usage in particular would be important in asymmetrical HMD-VR devices, such as the Vive Pro which trails the power and data cable down the left side of the HMD. The use of single-ear referencing with asymmetrically designed

HMD-VR may not accurately remove the electrical noise across the scalp, resulting in distorted waveforms.

6.2.7) Maze Task Limitations and Future Directions

The mazes developed for the spatial navigation task were designed to increase in difficulty within displays to measure three distinct levels of WML during active learning and recall navigation. The results found that whilst passively navigating a maze resulted in a lower WML compared to active learning or recall navigation, there was no differences in the neurophysiological measures of WML between maze difficulties. A total of two participants recruited for the spatial navigation task were excluded for not passing a run in a maze recall block, one for failing an 8-turn maze, and the other for a 12-turn maze. Considering the high success rate, and that previous studies comparing between maze difficulties have reported behavioural and EEG measures of WML being different between maze difficulties and between learning & recall routes [369,370], it is unlikely the lack of difference in WML found resulted from an overloaded working memory system in all conditions. Instead, a more likely explanation is that the presented mazes were similarly ‘easy’ for most participants.

Future studies should therefore ensure the gap in difficulty between simpler and more complex mazes is sufficiently wide enough to evoke different levels of WML. In the spatial navigation task, the 4 and 12 turns did not result in significantly distinct WMLs, so increasing the maximum number of turns to 16 may be more suitable. An alternative solution would be to increase the number of potential directions at each junction to 3, similar to that used by Hsieh et al. [414]. By increasing the number of potential directions at each junction, the possible number of turn sequences increases exponentially, and the chances of a participant guessing the correct direction at a junction is reduced.

A second factor that may have contributed to the relatively low difficulty of the higher-turn mazes is the repetition of the first 4 and 8 turns of the maze in subsequent maze

lengths. To comply with the limitations imposed by the Hammer editor and Garry's Mod, mazes were repeated between difficulties by increasing the amount of the complete maze participants explored. The improved recall of the repeated turn sequences could be similar to the 'Hebb repetition effect' reported in repeated exposure to lists of items, where repetition of early items in the list improves recall of later items [436]. It is therefore recommended that future studies utilise completely novel mazes between difficulties, and by extension use specialised software capable of developing mazes in larger and more complex VEs useable in DB-VR and HMD-VR.

6.2.8) Small Sample Sizes

A criticism that can be levelled against both experiments conducted in this thesis is the small sample sizes utilised compared to non-EEG research. Despite conforming to standard EEG practice for number of participants initially recruited, both studies could be argued to comprise of small sample sizes which may limit the power and generalisability of the conclusions reached [183,249]. Specifically, it is possible that the small sample size used hindered the statistical analysis conducted, and thus potentially introduced a source of bias or confound to the data [249]. This is particularly relevant for the spatial navigation experiment, where only 14 datasets were included in the final analysis, which likely contributed to the lack of significant findings between difficulties levels and between display conditions. Whilst the mental arithmetic tasks reached the average number of participants in an EEG study, previous spatial navigation tasks have reported recruiting 25 or more participants [370,385]. It is therefore suggested that future experiments conducted using EEG, particularly when combined with HMD-VR, utilise larger sample sizes to avoid biasing the data. This is especially important when considering the relative infancy of the combined EEG and HMD-VR literature, where the relatively small number of comparisons between other similar experiments can make identifying anomalous results more difficult.

This is not to suggest the limited EEG results found in the spatial navigation experiment resulted only from the small sample size used. Previous studies using

EEG have reported finding significant differences when using similar or lower numbers of participants, for example between 12-16 participants in spatial navigation experiments [70,371,429]. It is therefore suggested that, whilst the low number of participants likely contributed to the limited findings, it is also unlikely to be the only reason that minimal significant findings were reported.

It is also unknown if using a larger sample size would change the outcomes of the preprocessing pipeline comparison conducted in Chapter 4. However, it is unlikely the overall conclusions would change as the same data was manipulated between conditions. The consistent data means that any differences must result from the preprocessing decisions rather than differences within or between participants. Moreover, other comparisons of EEG preprocessing steps, such as Tanner et al. [310], also have used 22 participants, making it difficult to tell if a different number of participants would affect the outcomes.

6.3) Conclusions

This methodological thesis has examined the utility of combined HMD-VR and EEG in the study of working memory. Evidence is provided that HMD-VR is suitable for use both in combination with EEG, and in the study of working memory processes. HS-HMD-VR has been successfully combined with EEG recording measures in two working memory tasks utilising different ERP and power spectral density measures of WML. WML was successfully manipulated within the arithmetic task as evidenced by the behavioural and ERP results. Moreover, a distinction between passive and active navigation was found in the theta/alpha ratio within the spatial navigation task. Artifacts associated with the use of HMD-VR were successfully minimised using standard EEG preprocessing parameters.

It is also found that HMD-VR does not inherently increase WML relative to other display methods across the systematic review and in a comparison of WML between HS-HMD-VR and DB-VR in a maze task. The finding that HMD-VR does not introduce unwanted extraneous load means the advantages of the immersive

display, such as removing external distractions and allowing users to focus on a presented task, can be applied in both research settings and within real-world applications. Whilst cybersickness was found to be a problem for some individuals, it did not prevent the successful completion of the experiments by the majority of participants, and methods for reducing the impact of cybersickness in future research have now been are identified. Ultimately, this thesis concludes that HMD-VR is a useful tool for neuroscience and working memory based research, and should be further explored to identify how to best take advantage of the display method.

References

1. Foreman N. Virtual Reality in Psychology. THEMES IN SCIENCE AND TECHNOLOGY EDUCATION Special Issue. 2006; 225–252.

2. Pan X, Hamilton AF de C. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. British Journal of Psychology. 2018;109: 395–417. doi:10.1111/bjop.12290

3. Church RM. SESSION I PRESIDENTIAL ADDRESS The influence of computers on psychological research: A case study. Behavior Research Methods & Instrumentation. 1983.

4. Aaronson D. Computer use in cognitive psychology. Behavior Research Methods, Instruments, & Computers. 1994.

5. Cipresso P, Giglioli IAC, Raya MA, Riva G. The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. Front Psychol. 2018;9: 1–20. doi:10.3389/fpsyg.2018.02086

6. Rupp M, Odette KL, Kozachuk J, Michaelis JR, Smither JA, McConnell DS. Investigating learning outcomes and subjective experiences in 360-degree videos. Comput Educ. 2019;128: 256–268. doi:10.1016/j.compedu.2018.09.015

7. Curtin J, Lozano D, JJB Allen. The psychophysiological laboratory. Handbook of emotion elicitation and assessment. 2007. pp. 398–425.

8. Dickter CL, Kieffaber PD. The EEG Laboratory. EEG Methods for the Psychological Sciences. SAGE Publications Ltd; 2014.

9. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences. 2009;364: 3549–3557. doi:10.1098/rstb.2009.0138

10. Slater M. Immersion and the illusion of presence in virtual reality. British Journal of Psychology. Wiley/Blackwell (10.1111); 2018. pp. 431–433. doi:10.1111/bjop.12305

11. Reggente N, Essoe JKY, Aghajan ZM, Tavakoli A V., McGuire JF, Suthana NA, et al. Enhancing the ecological validity of fMRI memory research using virtual reality. Front Neurosci. 2018;12: 1–9. doi:10.3389/fnins.2018.00408

12. Hofmann SM, Klotzsche F, Mariola A, Nikulin V, Villringer A, Gaebler M. Decoding subjective emotional arousal from EEG during an immersive virtual reality experience. 2021;10. doi:10.7554/eLife

13. Wilson CJ, Soranzo A. The Use of Virtual Reality in Psychology: A Case Study in Visual Perception. Comput Math Methods Med. 2015;2015: 1–7. doi:10.1155/2015/151702

14. McCreery MP, Schrader PG, Krach SK, Boone R. A sense of self: The role of presence in virtual environments. 2013 [cited 25 Feb 2024]. doi:10.1016/j.chb.2013.02.002

15. Miller HL, Bugnariu NL. Level of Immersion in Virtual Environments Impacts the Ability to Assess and Teach Social Skills in Autism Spectrum Disorder. 2016;19: 246–256. doi:10.1089/cyber.2014.0682

16. Makransky G, Terkildsen TS, Mayer RE. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learn Instr. 2019;60: 225–236. doi:10.1016/j.learninstruc.2017.12.007

17. Schroeder R. Being There Together and the Future of Connected Presence. Presence: Teleoperators and Virtual Environments. 2006;15: 438–454. doi:10.1162/PRES.15.4.438

18. Loomis JM, Blascovich JJ, Beall AC. Immersive virtual environment technology as a basic research tool in psychology. Behavior Research Methods, Instruments, & Computers. 1999;31: 557–564.

19. Slater M, Sanchez-Vives M V. Enhancing our lives with immersive virtual reality. Frontiers Robotics AI. 2016;3. doi:10.3389/FROBT.2016.00074/ABSTRACT

20. Joseph A, Browning MHEM, Jiang S. Using Immersive Virtual Environments (IVEs) to Conduct Environmental Design Research: A Primer and Decision

Framework. Health Environments Research and Design Journal. 2020;13: 11–25. doi:10.1177/1937586720924787/FORMAT/EPUB

21. Rizzo AA, Bowerly T, Galen Buckwalter J, Mitura R, Parsons TD. A Virtual Reality Scenario for All Seasons: The Virtual Classroom. 2006 [cited 25 Feb 2024]. Available: www.ict.usc.edu.

22. Hasenbein L, Stark P, Trautwein U, Muller Queiroz AC, Bailenson J, Hahn J-U, et al. Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom. 2022 [cited 25 Feb 2024]. doi:10.1016/j.chb.2022.107282

23. Mapala T, Warmelink L, Linkenauger SA. Jumping the gun: Faster response latencies to deceptive questions in a realistic scenario. Psychon Bull Rev. 2017;24: 1350–1358. doi:10.3758/s13423-016-1218-z

24. Sobhani A, Farooq B. Impact of smartphone distraction on pedestrians' crossing behaviour: An application of head-mounted immersive virtual reality. Transp Res Part F Traffic Psychol Behav. 2018;58: 228–241. doi:10.1016/J.TRF.2018.06.020

25. Peterson SM, Furuichi E, Ferris DP. Effects of virtual reality high heights exposure during beam-walking on physiological stress and cognitive loading. PLoS One. 2018;13: 1–17. doi:10.1371/journal.pone.0200306

26. Clements JM, Kopper R, Zielinski DJ, Rao H, Sommer MA, Kirsch E, et al. Neurophysiology of Visual-Motor Learning during a Simulated Marksmanship Task in Immersive Virtual Reality. 25th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018 - Proceedings. 2018;9: 451–458. doi:10.1109/VR.2018.8446068

27. Nunes de Vasconcelos G, Malard ML, van Stralen M, Campomori M, Canavezzi de Abreu S, Lobosco T, et al. Do we still need CAVEs? 2020;3: 133–141. doi:10.5151/proceedings-ecaadesigradi2019_474

28. Mazuryk T, Gervautz M. Virtual Reality - History, Applications, Technology and Future. Virtual Real. 1996;72.

29. Boletsis C. The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. Multimodal Technologies and Interaction. 2017;1: 1–17. doi:10.3390/mti1040024

30. Krompiec P, Park K. Enhanced player interaction using motion controllers for first-person shooting games in virtual reality. IEEE Access. 2019;7: 124548–124557. doi:10.1109/ACCESS.2019.2937937

31. Sportillo D, Paljic A, Ojeda L. Get ready for automated driving using Virtual Reality Get Ready for Automated Driving using Virtual Reality. 2018. doi:10.1016/j.aap.2018.06.003

32. Oberhauser M, Dreyer D, Braunstingl R, Koglbauer I. What's real about virtual reality flight simulation? Comparing the fidelity of a virtual reality with a conventional flight simulation environment. Aviation Psychology and Applied Human Factors. 2018;8: 22–34. doi:http://dx.doi.org/10.1027/2192-0923/a000134

33. Amos E. Image of an Oculus Rift CV1 Headset From The Front. 2017 [cited 17 Jun 2021]. Available: https://en.wikipedia.org/wiki/Oculus_Rift_CV1#/media/File:Oculus-Rift-CV1-Headset-Front.jpg

34. Kwon C. Verification of the possibility and effectiveness of experiential learning using HMD-based immersive VR technologies. Virtual Real. 2019;23: 101–118. doi:10.1007/s10055-018-0364-1

35. HTC Corporation. HTC Vive Pro. 2018.

36. Cattan G, Andreev A, Mendoza C, Congedo M. The Impact of Passive Head-Mounted Virtual Reality Devices on the Quality of EEG Signals. VRIPHYS 2018 - 14th Workshop on Virtual Reality Interactions and Physical Simulations, VRIPHYS 2018. 2018; 21–27. doi:10.2312/VRIPHYS.20181064

37. Oculus VR. Oculus Quest. 2019.

38. Amos E. Image of an Google Cardboard From the Back. 2015 [cited 17 Jun 2021]. Available:

https://en.wikipedia.org/wiki/Google_Cardboard#/media/File:Google-Cardboard.jpg

39. Martirosov S, Bureš M, Tomáš Zítka ·. Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality. Virtual Real. 2022;26: 15–32. doi:10.1007/s10055-021-00507-4

40. Cruz-Neira C, Sandin DJ, DeFanti TA. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. Proc ACM SIGGRAPH 93 Conf Comput Graphics. 1993; 135–142.

41. James KH, Humphrey GK, Vilis T, Corrie B, Baddour R, Goodale MA. “Active” and “passive” learning of three-dimensional object structure within an immersive virtual reality environment. Behavior Research Methods, Instruments & Computers. 2002;34: 383–390. Available: Keith@uwo.ca

42. Molina G, Gimeno J, Portalés C, Casas S. A comparative analysis of two immersive virtual reality systems in the integration and visualization of natural hand interaction. Multimed Tools Appl. 2022;81: 7733–7758. doi:10.1007/S11042-021-11760-9/TABLES/16

43. Rauter G, Sigrist R, Koch C, Crivelli F, Van Raai M, Riener R, et al. Transfer of Complex Skill Learning from Virtual to Real Rowing. PLoS One. 2013;8: e82145. doi:10.1371/JOURNAL.PONE.0082145

44. Tauscher JP, Schottky FW, Grogorick S, Bittner PM, Mustafa M, Magnor M. Immersive EEG: Evaluating electroencephalography in virtual reality. 26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings. 2019; 1794–1800. doi:10.1109/VR.2019.8797858

45. Luck SJ. An introduction to the event-related potential technique. 2nd Edition. MIT Press; 2014.

46. Collura TF. History and evolution of electroencephalographic instruments and techniques. J Clin Neurophysiol. 1993;10: 476–504. doi:10.1097/00004691-199310000-00007

47. La Vaque TJ. Journal of Neurotherapy: Investigations in Neuromodulation, Neurofeedback and Applied Neuroscience The History of EEG Hans Berger. 1999 [cited 25 Feb 2024]. doi:10.1300/J184v03n02_01

48. Louis EK St., Frey LC, Britton JW, Frey LC, Hopp JL, Korb P, et al. Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants. Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants. 2016 [cited 25 Feb 2024]. Available: https://www.ncbi.nlm.nih.gov/books/NBK390354/

49. Dickter CL, Kieffaber PD. From Cortex to Computer: The Principles of Recording EEG. EEG Methods for the Psychological Sciences. SAGE Publications Ltd; 2014.

50. Tallgren P, Vanhatalo S, Kaila K, Voipio J. Evaluation of commercially available electrodes and gels for recording of slow EEG potentials. Clinical Neurophysiology. 2005;116: 799–806. doi:10.1016/J.CLINPH.2004.10.001

51. Klem G, Lüders HO, Jasper HH, Elger C. The ten-twenty electrode system of the International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol Suppl. 1958;52: 3–6.

52. Diagram of the International 10-20 System for EEG. 2010 [cited 5 Aug 2019]. Available: https://upload.wikimedia.org/wikipedia/commons/thumb/7/70/21_electrodes_of_International_10-20_system_for_EEG.svg/1024px-21_electrodes_of_International_10-20_system_for_EEG.svg.png

53. Jackson AF, Bolger DJ. The neurophysiological bases of EEG and EEG measurement: A review for the rest of us. Psychophysiology. 2014;51: 1061–1071. doi:10.1111/PSYP.12283

54. Teplan M. FUNDAMENTALS OF EEG MEASUREMENT. MEASUREMENT SCIENCE REVIEW. 2002;2.

55. Peirce J, Gray JR, Simpson S, MacAskill M, Höchenberger R, Sogo H, et al. PsychoPy2: Experiments in behavior made easy. Behav Res Methods. 2019;51: 195–203. doi:10.3758/S13428-018-01193-Y/FIGURES/3

56. Hertweck S, Weber D, Alwanni H, Unruh F, Fischbach M, Latoschik ME, et al. Brain activity in virtual reality: Assessing signal quality of high-resolution EEG while using head-mounted displays. 26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings. 2019; 970–971. doi:10.1109/VR.2019.8798369

57. Singh V, Veer K, Sharma R, Kumar S. Comparative study of FIR and IIR filters for the removal of 50 Hz noise from EEG signal. Int J Biomed Eng Technol. 2016;22: 250–257. doi:10.1504/IJBET.2016.079488

58. Hu L, Zhang Z. EEG Signal Processing and Feature Extraction. Springer; 2019.

59. Luck SJ, Emily S. Kappenman. The Oxford handbook of event-related potential components. Luck SJ, Kappenman ES, editors. Oxford university press; 2011.

60. Bayliss JD, Ballard DH. A virtual reality testbed for brain-computer interface research. IEEE Transactions on Rehabilitation Engineering. 2000;8: 188–190. doi:10.1109/86.847811

61. Cho BH, Ku J, Jang DP, Kim S, Lee YH, Kim IY, et al. The effect of virtual reality cognitive training for attention enhancement. Cyberpsychology and Behavior. 2002;5: 129–137. doi:10.1089/109493102753770516

62. Plank M, Snider J, Kaestner E, Halgren E, Poizner H. EEG correlates of unsupervised spatial learning in immersive, large-scale virtual environments. International IEEE/EMBS Conference on Neural Engineering, NER. 2013; 1346–1349. doi:10.1109/NER.2013.6696191

63. Tromp J, Peeters D, Meyer AS, Hagoort P. The combined use of virtual reality and EEG to study language processing in naturalistic environments. Behav Res Methods. 2018;50: 862–869. doi:10.3758/s13428-017-0911-9

64. Cattan GH, Andreev A, Mendoza C, Congedo M. A Comparison of Mobile VR Display Running on an Ordinary Smartphone with Standard PC Display for P300-BCI Stimulus Presentation. IEEE Trans Games. 2021;13: 68–77. doi:10.1109/TG.2019.2957963

65. Singh AK, Chen HT, Cheng YF, King JT, Ko LW, Gramann K, et al. Visual Appearance Modulates Prediction Error in Virtual Reality. IEEE Access. 2018;6: 24617–24624. doi:10.1109/ACCESS.2018.2832089

66. Rupp G, Berka C, Meghdadi A, McConnell MC, Storm M, Ramsøy TZ, et al. EEG Acquisition During the VR Administration of Resting State, Attention, and Image Recognition Tasks: A Feasibility Study. Communications in Computer and Information Science. 2019;1033: 250–258. doi:10.1007/978-3-030-23528-4_35/FIGURES/9

67. Hyun K-Y, Lee G-H. Analysis of Change of Event Related Potential in Escape Test using Virtual Reality Technology. Biomedical Science Letters. 2019;25: 139–148. doi:10.15616/BSL.2019.25.2.139

68. Meltzer JA, Zaveri HP, Goncharova II, Distasio MM, Papademetris X, Spencer SS, et al. Effects of Working Memory Load on Oscillatory Power in Human Intracranial EEG. Cerebral Cortex August. 2008;18: 1843–1855. doi:10.1093/cercor/bhm213

69. De Smedt B, Grabner RH, Studer B. Oscillatory EEG correlates of arithmetic strategy use in addition and subtraction. Exp Brain Res. 2009;195: 635–642. doi:10.1007/s00221-009-1839-9

70. Bischof WF, Boulanger P. Spatial Navigation in Virtual Reality Environments: An EEG Analysis. CyberPsychology & Behavior. 2003;6: 487–495. doi:10.1089/109493103769710514

71. Gramann K, Hohlefeld FU, Gehrke L, Klug M. Human cortical dynamics during full-body heading changes. Scientific Reports 2021 11:1. 2021;11: 1–12. doi:10.1038/s41598-021-97749-8

72. Celikcan U. Detection and Mitigation of Cybersickness via EEG-Based Visual Comfort Improvement. 3rd International Symposium on Multidisciplinary

Studies and Innovative Technologies, ISMSIT 2019 - Proceedings. 2019. doi:10.1109/ISMSIT.2019.8932870

73. Oh SH, Keun Whangbo T. Study on relieving VR contents user's fatigue degree using aroma by measuring EEG. 9th International Conference on Information and Communication Technology Convergence: ICT Convergence Powered by Smart Intelligence, ICTC 2018. 2018; 568–570. doi:10.1109/ICTC.2018.8539479

74. Bridgeman B, Lennon M Lou, Jackenthal A. Effects of Screen Size, Screen Resolution, and Display Rate on Computer-Based Test Performance. Applied Measurement in Education. 2003;16: 191–205. doi:10.1207/S15324818AME1603_2

75. Hancock PA, Sawyer BD, Stafford S. The effects of display size on performance. Ergonomics. 2015;58: 337–354. doi:10.1080/00140139.2014.973914

76. Hitch GJ, Baddeley AD. Verbal Reasoning and Working Memory. https://doi.org/101080/14640747608400587. 1976;28: 603–621. doi:10.1080/14640747608400587

77. Baddeley A. Working Memory Alan Baddeley. Science (1979). 1992;255: 556–559. doi:10.1126/science.1736359

78. Baddeley A. The episodic buffer: A new component of working memory? Trends Cogn Sci. 2000;4: 417–423. doi:10.1016/S1364-6613(00)01538-2

79. Baddeley A. Working memory: Theories, models, and controversies. Annu Rev Psychol. 2012;63: 1–29. doi:10.1146/ANNUREV-PSYCH-120710-100422

80. Shiffrin RM, Atkinson RC. Storage and retrieval processes in long-term memory. Psychol Rev. 1969;76: 179–193. doi:10.1037/H0027277

81. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive Architecture and Instructional Design. Educ Psychol Rev. 1998;10: 251–296. doi:10.1023/A:1022193728205

82. Cowan N. An Embedded-Processes Model of Working Memory. Models of Working Memory. 1999; 62–101. doi:10.1017/CBO9781139174909.006

83. Timarová S. Working memory and simultaneous interpreting. Translation and its others: selected papers of the CETRA research seminar in translation studies 2007. 2008;KULeuven; Leuven. Available: https://lirias.kuleuven.be/retrieve/214722

84. Adams EJ, Nguyen AT, Cowan N. Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose. Language, Speech, and Hearing Services in Schools. American Speech-Language-Hearing Association; 2018. pp. 340–355. doi:10.1044/2018_LSHSS-17-0114

85. Cowan N, Morey CC, Naveh-Benjamin M. An Embedded-Processes Approach to Working Memory: How Is It Distinct From Other Approaches, and to What Ends? Working Memory: State of the Science. Oxford University Press; 2020. pp. 44–84. doi:10.1093/oso/9780198842286.003.0003

86. Slana Ozimič A. Working Memory from the Perspective of the Multicomponent Model and Embedded-Processes Model. Interdisciplinary Description of Complex Systems : INDECS. 2020;18: 515–523. doi:10.7906/INDECS.18.4.2

87. Soemer A. The Multicomponent Working Memory Model, Attention, and Long-term Memory in Multimedia Learning: A Comment on Schweppe and Rummer (2014). Educ Psychol Rev. 2016;28: 197–200. doi:10.1007/S10648-015-9303-9/METRICS

88. Adams E, Forsberg A, Cowan N. The Embedded-Processes Model and Language Use. The Cambridge handbook of working memory and language. Submitted for publication: Cambridge University Press; 2021. Available: https://www.researchgate.net/publication/359013027

89. Cowan N. Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Constraints Within the Human Information-Processing System. Psychol Bull. 1988;104: 163–191. doi:10.1037/0033-2909.104.2.163

90. Miller GA. The Magical Number Seven, Plus or Minus Two. 1956.

91. Cowan N. Working Memory Underpins Cognitive Development, Learning, and Education. Educ Psychol Rev. 2014;26: 197. doi:10.1007/S10648-013-9246-Y

92. Lemaire P, Michel F, Fayol M. The Role of Working Memory Resources in Simple Cognitive Arithmetic The Role of Working Memory Resources in Simple Cognitive Arithmetic 74 WORKING MEMORY RESOURCES AND ARITHMETIC. Article in European Journal of Cognitive Psychology. 1996;8: 73–103. doi:10.1080/095414496383211

93. Berry AS, Zanto TP, Rutman AM, Clapp WC, Gazzaley A. Practice-related improvement in working memory is modulated by changes in processing external interference. J Neurophysiol. 2009;102: 1779–1789. doi:10.1152/JN.00179.2009/ASSET/IMAGES/LARGE/Z9K0090996780006.JPEG

94. Sweller J. Cognitive Load During Problem Solving: Effects on Learning. Cogn Sci. 1988;12: 257–285.

95. Sweller J, Chandler P. Evidence for Cognitive Load Theory Linked references are available on JSTOR for this article : Evidence for Cognitive Load Theory. 1991;8: 351–362.

96. Sweller J, Ayres P, Kalyuga S. Measuring Cognitive Load. Cognitive Load Theory. 2011; 71–85. doi:10.1007/978-1-4419-8126-4_6

97. Sweller J. Element interactivity and intrinsic, extraneous, and germane cognitive load. Educ Psychol Rev. 2010;22: 123–138. doi:10.1007/s10648-010-9128-5

98. Paas F, Renkl A, Sweller J. Cognitive Load Theory and Instructional Design: Recent Developments. Educ Psychol. 2003;38: 1–4. doi:10.1207/S15326985EP3801_1

99. Moreno R, Park B. Cognitive load theory: Historical development and relation to other theories. Cognitive Load Theory. 2010; 7–28. doi:10.1017/CBO9780511844744.003

100. Ayres P, Paas F. Cognitive Load Theory : New Directions and Challenges. 2012;832: 827–832.

101. Sweller J. Cognitive load theory, learning difficulty, and instructional design. Learn Instr. 1994;4: 295–312. doi:10.1016/0959-4752(94)90003-5

102. Van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ. 2010;44: 85–93. doi:10.1111/J.1365-2923.2009.03498.X

103. Mayer RE, Moreno R. Nine Ways to Reduce Cognitive Load in Multimedia Learning. Educ Psychol. 2003;38: 43–52. doi:10.1207/S15326985EP3801_6

104. DeLeeuw KE, Mayer RE. A Comparison of Three Measures of Cognitive Load: Evidence for Separable Measures of Intrinsic, Extraneous, and Germane Load. J Educ Psychol. 2008;100: 223–234. doi:10.1037/0022-0663.100.1.223

105. Parong J, Mayer R. Learning science in immersive virtual reality. J Educ Psychol. 2018;110: 785. Available: http://psycnet.apa.org/record/2018-03101-001

106. Blasiman RN, Larabee D, Fabry D. Distracted students: A comparison of multiple types of distractions on learning in online lectures. Scholarsh Teach Learn Psychol. 2018;4: 222–230. doi:10.1037/STL0000122

107. Martin S. Educational Research and Evaluation An International Journal on Theory and Practice Measuring cognitive load and cognition: metrics for technology-enhanced learning. 2015 [cited 27 Sep 2022]. doi:10.1080/13803611.2014.997140

108. Antonenko PD, Keil Andreas. Assessing working memory dynamics with electroencephalography. In: Zhang, editor. Cognitive load measurement and application: A theoretical framework for meaningful research and practice. 2017. pp. 93–111.

109. Zhang L, Wade J, Bian D, Fan J, Swanson A, Weitlauf A, et al. Cognitive Load Measurement in a Virtual Reality-based Driving System for Autism Intervention. IEEE Trans Affect Comput. 2017;8: 176–189. doi:10.1109/TAFFC.2016.2582490

110. Brünken R, Plass JL, Leutner D. Direct measurement of cognitive load in multimedia learning. Educ Psychol. 2003;38: 53–61. doi:10.1207/S15326985EP3801_7

111. Korbach A, Brünken R, Park B. Measurement of cognitive load in multimedia learning: a comparison of different objective measures. Instr Sci. 2017;45: 515–536. doi:10.1007/S11251-017-9413-5/TABLES/7

112. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology. 1988;52: 139–183. doi:10.1016/S0166-4115(08)62386-9

113. Hart SG. NASA-task load index (NASA-TLX); 20 years later. Proceedings of the Human Factors and Ergonomics Society. 2006; 904–908.

114. Nourbakhsh N, Wang Y, Chen F, Gsr C. Features for Cognitive Load Classifica-tion. 2013; 159–166. doi:10.1007/978-3-642-40483-2_11ï

115. Mitra R, Mcneal KS, Bondell HD. Pupillary response to complex interdependent tasks: A cognitive-load theory perspective. Behav Res Methods. 2017;49: 1905–1919. doi:10.3758/s13428-016-0833-y

116. Hossain D, Salimullah SM, Mahmudi R, Hasan SMN, Kabir E, Chowdhury AN, et al. Cognitive Load Measurement Using Galvanic Skin Response for Listening Tasks. 2019 4th International Conference on Electrical Information and Communication Technology, EICT 2019. 2019. doi:10.1109/EICT48899.2019.9068854

117. Emch M, von Bastian CC, Koch K. Neural correlates of verbal working memory: An fMRI meta-analysis. Front Hum Neurosci. 2019;13: 447322. doi:10.3389/FNHUM.2019.00180/BIBTEX

118. Fishburn FA, Norr ME, Medvedev A V., Vaidya CJ. Sensitivity of fNIRS to cognitive state and load. Front Hum Neurosci. 2014;8: 73786. doi:10.3389/FNHUM.2014.00076/ABSTRACT

119. Dong D, Wong LKF, Luo Z. Assessment of Prospective Memory using fNIRS in Immersive Virtual Reality Environment. J Behav Brain Sci. 2017;07: 247–258. doi:10.4236/jbbs.2017.76018

120. Yaple Z, Arsalidou M. N-back Working Memory Task: Meta-analysis of Normative fMRI Studies With Children. Child Dev. 2018;89: 2010–2022. doi:10.1111/CDEV.13080

121. Antonenko P, Paas F, Grabner R. Using Electroencephalography to Measure Cognitive Load. 2010; 425–438. doi:10.1007/s10648-010-9130-y

122. Chikhi S, Matton N, Blanchet S. EEG power spectral measures of cognitive workload: A meta-analysis. Psychophysiology. 2022;59: e14009. doi:10.1111/PSYP.14009

123. Gerjets P, Walter C, Rosenstiel W, Bogdan M, Zander TO. Cognitive state monitoring and the design of adaptive instruction in digital environments: Lessons learned from cognitive workload assessment using a passive brain-computer interface approach. Front Neurosci. 2014;8: 86703. doi:10.3389/FNINS.2014.00385/BIBTEX

124. Käthner I, Wriessnegger SC, Müller-Putz GR, Kübler A, Halder S. Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain-computer interface. Biol Psychol. 2014;102: 118–129. doi:10.1016/j.biopsycho.2014.07.014

125. Britton JW, Frey LC, Hopp JL, Korb P, Koubeissi MZ, Lievens WE, et al. Electroencephalography: An introductory Text and Atlas. Louis EKS, Frey LC, editors. Chicago, IL: American Epilepsy Society; 2016. doi:http://dx.doi.org/10.5698/978-0-9979756-0-4

126. Khader PH, Jost K, Ranganath C, Rösler F. Theta and alpha oscillations during working-memory maintenance predict successful long-term memory encoding. Neurosci Lett. 2010;468: 339–343. doi:10.1016/J.NEULET.2009.11.028

127. Schapkin SA, Raggatz J, Hillmert M, Böckelmann I. EEG correlates of cognitive load in a multiple choice reaction task. Acta Neurobiol Exp (Wars). 2020;80: 76–89. doi:10.21307/ane-2020-008

128. Antonenko PD, Niederhauser DS. Computers in Human Behavior The influence of leads on cognitive load and learning in a hypertext environment. Comput Human Behav. 2010;26: 140–150. doi:10.1016/j.chb.2009.10.014

129. Dan A, Reiner M. EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. International Journal of Psychophysiology. 2017;122: 75–84. doi:10.1016/j.ijpsycho.2016.08.013

130. Lee H. Measuring cognitive load with electroencephalography and self-report: focus on the effect of English-medium learning for Korean students. Educ Psychol (Lond). 2014;34: 838–848. doi:10.1080/01443410.2013.860217

131. Muluh ET. A Review of Event-Related Potential (Erp) Components Employed in Mental Arithmetic Processing Studies. J Neurol Neurophysiol. 2011;01. doi:10.4172/2155-9562.S6-001

132. Causse M, Fabre E, Giraudet L, Gonzalez M, Peysakhovich V. EEG/ERP as a Measure of Mental Workload in a Simple Piloting Task. Procedia Manuf. 2015;3: 5230–5236. doi:10.1016/j.promfg.2015.07.594

133. Polich J. Neuropsychology of P300. The Oxford Handbook of Event-Related Potential Components. 2011 [cited 11 Feb 2024]. doi:10.1093/OXFORDHB/9780195374148.013.0089

134. Scharinger C, Soutschek A, Schubert T, Gerjets P. Comparison of the working memory load in N-back and working memory span tasks by means of EEG frequency band power and P300 amplitude. Front Hum Neurosci. 2017;11: 212526. doi:10.3389/FNHUM.2017.00006/BIBTEX

135. Morgan HM, Klein C, Boehm SG, Shapiro KL, Linden DEJ. Working memory load for faces modulates P300, N170, and N250r. J Cogn Neurosci. 2008;20: 989. doi:10.1162/JOCN.2008.20072

136. Macnamara A, Schmidt J, Zelinsky GJ, Hajcak G. Electrocortical and ocular indices of attention to fearful and neutral faces presented under high and low working memory load. Biol Psychol. 2012;91: 349–356. doi:10.1016/j.biopsycho.2012.08.005

137. Chai WJ, Abd Hamid AI, Abdullah JM. Working memory from the psychological and neurosciences perspectives: A review. Front Psychol. 2018;9: 1–16. doi:10.3389/fpsyg.2018.00401

138. Concannon BJ, Esmail S, Roduta Roberts M. Head-Mounted Display Virtual Reality in Post-secondary Education and Skill Training. Front Educ (Lausanne). 2019;4: 456078. doi:10.3389/FEDUC.2019.00080/BIBTEX

139. Fabris CP, Rathner JA, Fong AY, Sevigny CP. Virtual Reality in Higher Education. International Journal of Innovation in Science and Mathematics Education. 2019;27: 69–80. doi:10.30722/IJISME.27.08.006

140. Radianti J, Majchrzak TA, Fromm J, Wohlgenannt I. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Comput Educ. 2020;147: 103778. doi:10.1016/j.compedu.2019.103778

141. Roettl J, Terlutter R. The same video game in 2D, 3D or virtual reality – How does technology impact game evaluation and brand placements? PLoS One. 2018;13: 1–25. doi:10.1371/journal.pone.0200724

142. Luong T, Martin N, Argelaguet Sanz F, Lécuyer A, Argelaguet F. Studying the Mental Effort in Virtual Versus Real Environments. 2019; 809–816. doi:10.1109/VR.2019.8798029ï

143. Jung J, Ahn YJ. Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study. Comput Animat Virtual Worlds. 2018;29: 1–10. doi:10.1002/cav.1812

144. Parmar D, Bertrand J, Babu S V., Madathil K, Zelaya M, Wang T, et al. A comparative evaluation of viewing metaphors on psychophysical skills education in an interactive virtual environment. Virtual Real. 2016;20: 141–157. doi:10.1007/s10055-016-0287-7

145. Roldán JJ, Peña-Tapia E, Martín-Barrio A, Olivares-Méndez MA, del Cerro J, Barrientos A. Multi-robot interfaces and operator situational awareness: Study of the impact of immersion and prediction. Sensors (Switzerland). 2017;17: 1–25. doi:10.3390/s17081720

146. Murcia-López M, Steed A. The effect of environmental features, self-avatar, and immersion on object location memory in virtual environments. Frontiers in ICT. 2016;3: 211015. doi:10.3389/FICT.2016.00024/BIBTEX

147. Jensen L, Konradsen F. A review of the use of virtual reality head-mounted displays in education and training. Educ Inf Technol (Dordr). 2018;23: 1515–1529. Available:

http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=130285744&site=ehost-live

148. Hamad A, Jia B. How Virtual Reality Technology Has Changed Our Lives: An Overview of the Current and Potential Applications and Limitations. Int J Environ Res Public Health. 2022;19. doi:10.3390/IJERPH191811278

149. Passig D, Tzuriel D, Eshel-Kedmi G. Improving children's cognitive modifiability by dynamic assessment in 3D Immersive Virtual Reality environments. Comput Educ. 2016;95: 296–308. doi:10.1016/J.COMPEDU.2016.01.009

150. Ray AB, Deb S. Smartphone Based Virtual Reality Systems in Classroom Teaching - A Study on the Effects of Learning Outcome. Proceedings - IEEE 8th International Conference on Technology for Education, T4E 2016. 2017; 68–71. doi:10.1109/T4E.2016.022

151. Webster R. Declarative knowledge acquisition in immersive virtual learning environments. Interactive Learning Environments. 2016;24: 1319–1333. doi:10.1080/10494820.2014.994533

152. Frederiksen JG, Sørensen SMD, Konge L, Svendsen MBS, Nobel-Jørgensen M, Bjerrum F, et al. Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. Surg Endosc. 2020;34: 1244–1252. doi:10.1007/s00464-019-06887-8

153. Kemeny A, George P, Mérienne F, Colombet F. New VR Navigation Techniques to Reduce Cybersickness. Electronic Imaging, The Engineering Reality of Virtual Reality. 2017; 48–53. doi:10.2352/ISSN.2470-1173.2017.3.ERVR-097

154. Cobb SVG, Nichols S, Ramsey A, Wilson JR. Virtual Reality-Induced Symptoms and Effects (VRISE). Presence: Teleoperators and Virtual Environments. 1999;8: 169–186. doi:10.1162/105474699566152

155. Sharples S, Cobb S, Moody A, Wilson JR. Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop

and projection display systems. 2007 [cited 19 Dec 2023]. doi:10.1016/j.displa.2007.09.005

156. Fernandes AS, Feiner SK. Combating VR sickness through subtle dynamic field-of-view modification. 2016 IEEE Symposium on 3D User Interfaces, 3DUI 2016 - Proceedings. 2016; 201–210. doi:10.1109/3DUI.2016.7460053

157. Llorach G, Evans A, Blat J. Simulator Sickness and Presence using HMDs: comparing use of a game controller and a position estimation system. Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology. 2014; 137–140. doi:10.1145/2671015.2671120

158. Keshavarz B, Hecht H, Zschutschke L. Intra-visual conflict in visually induced motion sickness. 2011. doi:10.1016/j.displa.2011.05.009

159. Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. Virtual Real. 2016;20: 101–125. doi:10.1007/s10055-016-0285-9

160. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. https://doi.org/101207/s15327108ijap0303_3. 1993;3: 203–220. doi:10.1207/S15327108IJAP0303_3

161. Polcar J, Horejsi P. Knowledge acquisition and cyber sickness: A comparison of VR devices in virtual tours. MM Science Journal. 2015;2015: 613–616. doi:10.17973/MMSJ.2015_06_201516

162. Alves Fernandes LM, Cruz Matos G, Azevedo D, Rodrigues Nunes R, Paredes H, Morgado L, et al. Exploring educational immersive videogames: an empirical study with a 3D multimodal interaction prototype. Behaviour and Information Technology. 2016;35: 907–918. doi:10.1080/0144929X.2016.1232754

163. Dorado JL, Figueroa PA. Ramps are better than stairs to reduce cybersickness in applications based on a HMD and a Gamepad. IEEE Symposium on 3D User Interfaces 2014, 3DUI 2014 - Proceedings. 2014; 47–50. doi:10.1109/3DUI.2014.6798841

164. Stanney K, Lawson BD, Rokers B, Dennison M, Fidopiastis C, Stoffregen T, et al. Identifying Causes of and Solutions for Cybersickness in Immersive Technology: Reformulation of a Research and Development Agenda. Int J Hum Comput Interact. 2020;36: 1783–1803. doi:10.1080/10447318.2020.1828535

165. Mai C, Hassib M, Königbauer R. Estimating visual discomfort in head-mounted displays using electroencephalography. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2017;10516 LNCS: 243–252. doi:10.1007/978-3-319-68059-0_15

166. Srivastava P, Rimzhim A, Vijay P, Singh S, Chandra S. Desktop VR Is Better Than Non-ambulatory HMD VR for Spatial Learning. Front Robot AI. 2019;6: 1–15. doi:10.3389/frobt.2019.00050

167. Harjunen VJ, Ahmed I, Jacucci G, Ravaja N, Spapé MM. Manipulating bodily presence affects cross-modal spatial attention: A virtual-reality-based ERP study. Front Hum Neurosci. 2017;11. doi:10.3389/fnhum.2017.00079

168. Israel K, Zerres C, Tscheulin DK. Reducing cybersickness: The role of wearing comfort and ease of use. Optics InfoBase Conference Papers. 2017;Part F66-F: 7–8. doi:10.1364/FIO.2017.JTu2A.113

169. Cho B-H, Kim S, Shin DI, Lee JH, Lee SM, Kim IY, et al. Neurofeedback Training with Virtual Reality for Inattention and Impulsiveness. CyberPsychology & Behavior. 2004;7: 519–526.

170. Jongbloed-Pereboom M, Janssen AJWM, Steenbergen B, Nijhuis-Van Der Sanden MWG. Motor learning and working memory in children born preterm: A systematic review. Neurosci Biobehav Rev. 2012;36: 1314–1330. doi:10.1016/j.neubiorev.2012.02.005

171. Spencer-Smith M, Klingberg T. Benefits of a Working Memory Training Program for Inattention in Daily Life: A Systematic Review and Meta-Analysis. 2015 [cited 31 Oct 2024]. doi:10.1371/journal.pone.0119522

172. Ferreira dos Santos L, Christ O, Mate K, Schmidt H, Krüger J, Dohle C. Movement visualisation in virtual reality rehabilitation of the lower limb: A systematic review. Biomed Eng Online. 2016;15: 75–88. doi:10.1186/s12938-016-0289-4

173. Burns CG, Fairclough SH. Use of auditory event-related potentials to measure immersion during a computer game. International Journal of Human Computer Studies. 2015;73: 107–114. doi:10.1016/j.ijhcs.2014.09.002

174. Dong D, Wong LKF, Luo Z. Assess BA10 activity in slide-based and immersive virtual reality prospective memory task using functional near-infrared spectroscopy (fNIRS). Applied Neuropsychology:Adult. 2019;26: 465–471. doi:10.1080/23279095.2018.1443104

175. Juliano JM, Spicer RP, Vourvopoulos A, Lefebvre S, Jann K, Ard T, et al. Embodiment is related to better performance on a brain–computer interface in immersive virtual reality: A pilot study. Sensors (Switzerland). 2020;20. doi:10.3390/s20041204

176. Käthner I, Kübler A, Halder S. Rapid P300 brain-computer interface communication with a head-mounted display. Front Neurosci. 2015;9: 1–13. doi:10.3389/fnins.2015.00207

177. Li G, Anguera JA, Javed S V., Khan MA, Wang G, Gazzaley A. Enhanced attention using head-mounted virtual reality. J Cogn Neurosci. 2020;32: 1438–1434. doi:10.1162/jocn_a_01560

178. Peterson SM, Rios E, Ferris DP. Transient visual perturbations boost short-term balance learning in virtual reality by modulating electrocortical activity. J Neurophysiol. 2018;120: 1998–2010. doi:10.1152/jn.00292.2018

179. Škola F, Liarokapis F. Embodied VR environment facilitates motor imagery brain–computer interface training. Computers and Graphics (Pergamon). 2018;75: 59–71. doi:10.1016/j.cag.2018.05.024

180. Sun R, Wu YJ, Cai Q. The effect of a virtual reality learning environment on learners' spatial ability. Virtual Real. 2018;23: 385–398. doi:10.1007/s10055-018-0355-2

181. Xu W, Liang H-N, Zhang Z, Baghaei N. Studying the Effect of Display Type and Viewing Perspective on User Experience in Virtual Reality Exergames. Games Health J. 2020. doi:10.1089/g4h.2019.0102

182. Yang X, Lin L, Cheng PY, Yang X, Ren Y, Huang YM. Examining creativity through a virtual reality support system. Educational Technology Research and Development. 2018;66: 1231–1254. doi:10.1007/s11423-018-9604-z

183. Larson MJ, Carbine KA. Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor. 2016 [cited 3 Oct 2024]. doi:10.1016/j.ijpsycho.2016.06.015

184. Yücel MA, Lühmann A v., Scholkmann F, Gervain J, Dan I, Ayaz H, et al. Best practices for fNIRS publications. Neurophotonics. 2021;8: 012101. doi:10.1117/1.NPH.8.1.012101

185. Oculus VR LLC. Oculus Best Practices. Oculus Documentation. 2017; 36. doi:10.2106/JBJS.H.00657

186. Takada M, Fukui Y, Matsuura Y, Sato M, Takada H. Peripheral viewing during exposure to a 2D/3D video clip: effects on the human body. Environ Health Prev Med. 2015;20: 79–89. doi:10.1007/S12199-014-0424-4/FIGURES/6

187. Zhang C. Investigation on Motion Sickness in Virtual Reality Environment from the Perspective of User Experience. Proceedings of 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education, ICISCAE 2020. 2020; 393–396. doi:10.1109/ICISCAE51034.2020.9236907

188. Kim S, Park S, Jeong H, Sung J. Study on response time measurement of distracted driving by virtual reality driving simulator. International Journal of Advanced Computer Research. 2019;9: 2277–7970. doi:10.19101/IJACR.MUL16002

189. Hoffman HG, Richards TL, Coda B, Bills AR, Blough D, Richards AL, et al. Modulation of thermal pain-related brain activity with virtual reality: Evidence from fMRI. Neuroreport. 2004;15: 1245–1248. doi:10.1097/01.wnr.0000127826.73576.91

190. Juliano JM, Liew SL. Transfer of motor skill between virtual reality viewed using a head-mounted display and conventional screen environments. J Neuroeng Rehabil. 2020;17: 48. doi:10.1186/s12984-020-00678-2

191. Grassini S, Laumann K. Are Modern Head-Mounted Displays Sexist? A Systematic Review on Gender Differences in HMD-Mediated Virtual Reality. Front Psychol. 2020;11. doi:10.3389/fpsyg.2020.01604

192. Gonzalez-Franco M, Abtahi P, Steed A. Individual differences in embodied distance estimation in virtual reality. 26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings. 2019; 941–943. doi:10.1109/VR.2019.8798348

193. Coxon M, Kelly N, Page S. Individual differences in virtual reality: Are spatial presence and spatial ability linked? Virtual Real. 2016;20: 203–212. doi:10.1007/s10055-016-0292-x

194. Davis S, Nesbitt K, Nalivaiko E. Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015). 2015;167: 3–14. doi:10.17973/MMSJ.2015

195. Bockelman P, Lingum D. Factors of cybersickness. Communications in Computer and Information Science. 2017;714: 3–8. doi:10.1007/978-3-319-58753-0_1

196. Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: A review. Front Psychol. 2019;10: 1–19. doi:10.3389/fpsyg.2019.00158

197. Kemeny A, Chardonnet J-R, Colombet F. Reducing Cybersickness. Getting Rid of Cybersickness. 2020; 93–132. doi:10.1007/978-3-030-59342-1_4

198. Demitriadou E, Stavroulia KE, Lanitis A. Comparative evaluation of virtual and augmented reality for teaching mathematics in primary education. Educ Inf Technol (Dordr). 2020;25: 381–401. doi:10.1007/s10639-019-09973-5

199. Halabi O, Bahameish MA, Al-Naimi LT, Al-Kaabi AK. Response Times for Auditory and Vibrotactile Directional Cues in Different Immersive Displays. Int J

Hum Comput Interact. 2019;35: 1578–1585. doi:10.1080/10447318.2018.1555743

200. Funk M, Kosch T, Schmidt A. Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016; 934–939. doi:10.1145/2971648.2971706

201. Werrlich S, Daniel A, Ginger A, Nguyen PA, Notni G. Comparing HMD-Based and Paper-Based Training. Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2018. 2018; 134–142. doi:10.1109/ISMAR.2018.00046

202. Huang KT, Ball C, Francis J, Ratan R, Boumis J, Fordham J. Augmented versus virtual reality in education: An exploratory study examining science knowledge retention when using augmented reality/virtual reality mobile applications. Cyberpsychol Behav Soc Netw. 2019;22: 105–110. doi:10.1089/cyber.2018.0150

203. Fishburn FA, Norr ME, Medvedev A V., Vaidya CJ. Sensitivity of fNIRS to cognitive state and load. Front Hum Neurosci. 2014;8: 1–11. doi:10.3389/fnhum.2014.00076

204. Sun Y, Ding Y, Jiang J, Duffy VG. Measuring mental workload using ERPs based on FIR, ICA, and MARA. Computer Systems Science and Engineering. 2022;41: 781–794. doi:10.32604/csse.2022.016387

205. Aksoy M, Ufodiama CE, Bateson AD, Martin S, Asghar AUR. A comparative experimental study of visual brain event-related potentials to a working memory task: virtual reality head-mounted display versus a desktop computer screen. Exp Brain Res. 2021;239: 3007–3022. doi:10.1007/S00221-021-06158-W/TABLES/6

206. Stolz C, Endres D, Mueller EM. Threat-conditioned contexts modulate the late positive potential to faces—A mobile EEG/virtual reality study. Psychophysiology. 2019;56: 15. doi:10.1111/psyp.13308

207. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ. 2017;10: 549–559. doi:10.1002/ase.1696

208. Chénéchal M Le, Chatel-Goldman J. HTC Vive Pro time performance benchmark for scientific research. 2018 [cited 11 Dec 2023]. Available: https://hal.science/hal-01934741

209. Cummings JJ, Bailenson JN. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. Media Psychol. 2016;19: 272–309. doi:10.1080/15213269.2015.1015740

210. Başar E, Başar-Eroglu C, Karakaş S, Schürmann M. Gamma, alpha, delta, and theta oscillations govern cognitive processes. International Journal of Psychophysiology. 2001;39: 241–248. doi:10.1016/S0167-8760(00)00145-8

211. Nunez PL, Srinivasan R. Electric Fields of the Brain The Neurophysics of EEG Second Edition. 2006.

212. Costa R, Gomes PV, Correia A, Marques A, Pereira J. The Influence of Brain Activity on the Interactive Process through Biofeedback Mechanisms in Virtual Reality Environments. Engineering Proceedings 2021, Vol 7, Page 15. 2021;7: 15. doi:10.3390/ENGPROC2021007015

213. Lee JT, Janaka Rajapakse RPC, Miyata K. EEG-based Evaluation on Intuitive Gesture Interaction in Virtual Environment. Proceedings - 2022 International Conference on Cyberworlds, CW 2022. 2022; 213–217. doi:10.1109/CW55638.2022.00050

214. Chang Z, Bai H, Zhang L, Gupta K, He W, Billinghurst M. The impact of virtual agents' multimodal communication on brain activity and cognitive load in Virtual Reality. Front Virtual Real. 2022;3. doi:10.3389/frvir.2022.995090

215. Lehnert G, Zimmer HD. Auditory and visual spatial working memory. Mem Cognit. 2006;34: 1080–1090. doi:10.3758/BF03193254/METRICS

216. Lehnert G, Zimmer HD. Modality and domain specific components in auditory and visual working memory tasks. Cogn Process. 2008;9: 53–61. doi:10.1007/S10339-007-0187-6/FIGURES/3

217. Klingner J, Tversky B, Hanrahan P. Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. Psychophysiology. 2011;48: 323–332. doi:10.1111/j.1469-8986.2010.01069.x

218. Albus P, Seufert T. The modality effect reverses in a virtual reality learning environment and influences cognitive load. Instr Sci. 2023;51: 545–570. doi:10.1007/S11251-022-09611-7/FIGURES/4

219. DeStefano D, Lefevre J-A. The role of working memory in mental arithmetic. European Journal of Cognitive Psychology. 2004;16: 353–386. doi:10.1080/09541440244000328

220. Hubber PJ, Gilmore C, Cragg L. The roles of the central executive and visuospatial storage in mental arithmetic: A comparison across strategies. Quarterly Journal of Experimental Psychology. 2014;67: 936–954. doi:10.1080/17470218.2013.838590/ASSET/IMAGES/LARGE/10.1080_17470218.2013.838590-FIG4.JPEG

221. Núñez-Peña IM, Honrubia-Serrano LM, Escera C. Problem size effect in additions and subtractions: an event-related potential study. Neurosci Lett. 2005;373: 21–25. doi:10.1016/j.neulet.2004.09.053

222. Van Beek L, Ghesquière P, De Smedt B, Lagae L. The arithmetic problem size effect in children: An event-related potential study. Front Hum Neurosci. 2014;8: 1–11. doi:10.3389/FNHUM.2014.00756/BIBTEX

223. Dickson DS, Federmeier KD. The language of arithmetic across the hemispheres: an event-related potential investigation. Brain Res. 2018;1662: 46. doi:10.1016/J.BRAINRES.2017.02.019

224. Moro C, Stromberga Z, Stirling A. Virtualisation devices for student learning: Comparison between desktop-based (Oculus Rift) and mobile-based (Gear VR) virtual reality in medical and health science education. Australasian Journal of Educational Technology. 2017;33: 1–10. doi:10.14742/AJET.3840

225. Stepan K, Zeiger J, Hanchuk S, Del Signore A, Shrivastava R, Govindaraj S, et al. Immersive virtual reality as a teaching tool for neuroanatomy. Int Forum Allergy Rhinol. 2017;7: 1006–1013. doi:10.1002/ALR.21986

226. Seyler DJ, Kirk EP, Ashcraft MH. Elementary Subtraction. J Exp Psychol Learn Mem Cogn. 2003;29: 1339–1352. doi:10.1037/0278-7393.29.6.1339

227. Spüler M, Walter C, Rosenstiel W, Gerjets P, Moeller K, Klein E. EEG-based prediction of cognitive workload induced by arithmetic: a step towards online adaptation in numerical learning. ZDM - Mathematics Education. 2016;48: 267–278. doi:10.1007/s11858-015-0754-8

228. Furst AJ, Hitch GJ. Separate roles for executive and phonological components of working memory in mental arithmetic. Mem Cognit. 2000;28: 774–782.

229. Chin ZY, Zhang X, Wang C, Ang KK. EEG-based discrimination of different cognitive workload levels from mental arithmetic. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018; 1984–1987. doi:10.1109/EMBC.2018.8512675

230. Ashcraft MH, Guillaume MM. Chapter 4 Mathematical Cognition and the Problem Size Effect. Psychology of Learning and Motivation - Advances in Research and Theory. 2009. pp. 121–151. doi:10.1016/S0079-7421(09)51004-3

231. Raghubar KP, Barnes MA, Hecht SA. Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. Learn Individ Differ. 2010;20: 110–122. doi:10.1016/J.LINDIF.2009.10.005

232. Blankenberger S. The arithmetic tie effect is mainly encoding-based. Cognition. 2001;82: B15–B24. doi:10.1016/S0010-0277(01)00140-8

233. Hitch GJ. The role of short-term working memory in mental arithmetic. Cogn Psychol. 1978;10: 302–323. doi:10.1016/0010-0285(78)90002-6

234. Imbo I, Duverne S, Lemaire P. Working Memory, Strategy Execution, and Strategy Selection in Mental Arithmetic. https://doi.org/101080/17470210600943419. 2007;60: 1246–1264. doi:10.1080/17470210600943419

235. Ullsperger P, Neumann U, Gille HG, Pietschmann M. P300 and anticipated task difficulty. International Journal of Psychophysiology. 1987;5: 145–149. doi:10.1016/0167-8760(87)90018-3

236. Iguchi Y, Hashimoto I. Sequential information processing during a mental arithmetic is reflected in the time course of event-related brain potentials. Clinical Neurophysiology. 2000;111: 204–213. doi:10.1016/S1388-2457(99)00244-8

237. Jasinski EC, Coch D. ERPs across arithmetic operations in a delayed answer verification task. Psychophysiology. 2012;49: 943–958. doi:10.1111/J.1469-8986.2012.01378.X

238. Kiefer M, Dehaene S. The Time Course of Parietal Activation in Single-digit Multiplication: Evidence from Event-related Potentials. Mathematical Cognition. 1997;3: 1–30. doi:10.1080/135467997387461

239. Dickson DS, Wicha NYY. P300 amplitude and latency reflect arithmetic skill: An ERP study of the problem size effect. 2019 [cited 5 Jun 2023]. doi:10.1016/j.biopsycho.2019.107745

240. Suárez-Pellicioni M, Núñez-Peña MI, Colomé A. Mathematical anxiety effects on simple arithmetic processing efficiency: An event-related potential study. Biol Psychol. 2013;94: 517–526. doi:10.1016/J.BIOPSYCHO.2013.09.012

241. Davis Moore R, Drollette ES, Scudder MR, Bharij A, Hillman CH. The influence of cardiorespiratory fitness on strategic, behavioral, and electrophysiological indices of arithmetic cognition in preadolescent children. Front Hum Neurosci. 2014;8: 77985. doi:10.3389/FNHUM.2014.00258/BIBTEX

242. Eimer M, Kiss M, Nicholas S. Response Profile of the Face-Sensitive N170 Component: A Rapid Adaptation Study. Cerebral Cortex. 2010;20: 2442–2452. doi:10.1093/cercor/bhp312

243. He WQ, Luo WB, He HM, Chen X, Zhang DJ. N170 effects during exact and approximate calculation tasks: An ERP study. Neuroreport. 2011;22: 437–441. doi:10.1097/WNR.0B013E32834702C1

244. Muluh ET, Vaughan CL, John LR. High resolution event-related potentials analysis of the arithmetic-operation effect in mental arithmetic. Clinical Neurophysiology. 2011;122: 518–529. doi:10.1016/j.clinph.2010.08.008

245. Niedeggen M, Rösler F. N400 Effects Reflect Activation Spread During Retrieval of Arithmetic Facts. https://doi.org/101111/1467-928000149. 1999;10: 271–276. doi:10.1111/1467-9280.00149

246. Núñez-Peña IM, Honrubia-Serrano LM. P600 related to rule violation in an arithmetic task. 2004 [cited 9 Mar 2023]. doi:10.1016/j.cogbrainres.2003.09.010

247. Domahs F, Domahs U, Schlesewsky M, Ratinckx E, Verguts T, Willmes K, et al. Neighborhood consistency in mental arithmetic: Behavioral and ERP evidence. Behavioral and Brain Functions. 2007;3: 1–13. doi:10.1186/1744-9081-3-66/FIGURES/3

248. Ku Y, Hong B, Gao X, Gao S. Spectra-temporal patterns underlying mental addition: An ERP and ERD/ERS study. Neurosci Lett. 2010;472: 5–10. doi:10.1016/j.neulet.2010.01.040

249. Clayson PE, Carbine KA, Baldwin SA, Larson MJ. Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. Psychophysiology. 2019;56. doi:10.1111/PSYP.13437

250. Oldfield RC. THE ASSESSMENT AND ANALYSIS OF HANDEDNESS: THE EDINBURGH INVENTORY. Neuropsychologia. Pergamon Press; 1971.

251. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. JAMA. 2013;310: 2191–2194. doi:10.1001/JAMA.2013.281053

252. Van Rossum G, Drake FL. Introduction to Python 3. 2009.

253. Durette PN. Google Text-to-Speech. Pierre Nicolas Durette; 2018.

254. MediaHuman. MediaHuman Audio Converter. https://www.mediahuman.com/audio-converter/; 2018.

255. Bateson AD, Asghar AUR. Development and Evaluation of a Smartphone-Based Electroencephalography (EEG) System. IEEE Access. 2021;9: 75650–75667. doi:10.1109/ACCESS.2021.3079992

256. Kong J, Wang Y, Shang H, Wang Y, Yang X, Zhuang D. Brain potentials during mental arithmetic-effects of problem difficulty on event-related brain potentials. Neurosci Lett. 1999;260: 169–172. doi:10.1016/S0304-3940(98)00974-4

257. Paas FGWC. Training strategies for attaining transfer of problem solving skill in statistics. A cognitiveload approach. 1992;84: 429–434. doi:10.1037/0022-0663.84.4.429

258. Paas F, Tuovinen JE, Tabbers H, Van Gerven PWM. Cognitive load measurement as a means to advance cognitive load theory. Educ Psychol. 2003;38: 63–71. doi:10.1207/S15326985EP3801_8

259. Sweller J, van Merriënboer JJG, Paas F. Cognitive Architecture and Instructional Design: 20 Years Later. Educ Psychol Rev. 2019;31: 261–292. doi:10.1007/S10648-019-09465-5/FIGURES/1

260. Ayres P. Using subjective measures to detect variations of intrinsic cognitive load within problems. Learn Instr. 2006;16: 389–400. doi:10.1016/j.learninstruc.2006.09.001

261. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004;134: 9–21. Available: http://www.sccn.ucsd.edu/eeglab/

262. The MathWorks Inc. MATLAB . Massachusetts: The MathWorks Inc. ; 2022.

263. Lopez-Calderon J, Luck SJ. ERPLAB: An open-source toolbox for the analysis of event-related potentials. Front Hum Neurosci. 2014;8: 75729. doi:10.3389/FNHUM.2014.00213/BIBTEX

264. IBM Corp. IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp; 2021.

265. Microsoft Corporation. Microsoft Excel. 2022.

266. Hess LJ, Greenlee ET. The Stress of Vigilance May Influence Reports of Simulator Sickness: A Factor Analysis of the SSSQ and SSQ in Auditory Vigilance Tasks. 2022. doi:10.1177/1071181322661446

267. Field A. Discovering Statistics using IBM SPSS Statistics. 4th ed. Sage; 2013.

268. Ashcraft MH, Kirk EP. The Relationships Among Working Memory, Math Anxiety, and Performance. J Exp Psychol Gen. 2001;130: 224–237. doi:10.1037/0096-3445.130.2.224

269. Rebsamen B, Kwok K, Penney TB. Evaluation Of Cognitive Workload From EEG During A Mental Arithmetic Task. http://dx.doi.org/101177/1071181311551279. 2011; 1342–1345. doi:10.1177/1071181311551279

270. Campbell JID, Zbrodoff NJ, Logan GD. Handbook of Mathematical Cognition What Everyone Finds Publication details. 2004. doi:10.4324/9780203998045.ch19

271. Somrak A, Humar I, Hossain MS, Alhamid MF, Hossain MA, Guna J. Estimating VR Sickness and user experience using different HMD technologies: An evaluation study. Future Generation Computer Systems. 2019;94: 302–316. doi:10.1016/J.FUTURE.2018.11.041

272. Pratt N, Willoughby A, Swick D. Effects of working memory load on visual selective attention: Behavioral and electrophysiological evidence. Front Hum Neurosci. 2011;5: 1–9. doi:10.3389/fnhum.2011.00057

273. Wang L, Xu G, Yang S, Song Y, Wei Y, Yan W. Research on event related potential elicited by number recognizing and arithmetic calculating. Proc of 2007 Joint Meet of the 6th Int Symp on Noninvasive Functional Source Imaging of the Brain and Heart and the Int Conf on Functional Biomedical Imaging, NFSI and ICFBI 2007. 2007; 247–250. doi:10.1109/NFSI-ICFBI.2007.4387742

274. Polich J. Updating P300: An Integrative Theory of P3a and P3b. Clinical Neurophysiology. 2007;118: 2128–1248. doi:10.1016/j.clinph.2007.04.019

275. Maurer U, Rossion B, McCandliss BD. Category specificity in early perception: Face and word N170 responses differ in both lateralization and habituation properties. Front Hum Neurosci. 2008;2: 18. doi:10.3389/NEURO.09.018.2008/BIBTEX

276. Aranda C, Madrid E, Tudela P, Ruz M. Category expectations: A differential modulation of the N170 potential for faces and words. Neuropsychologia. 2010;48: 4038–4045. doi:10.1016/J.NEUROPSYCHOLOGIA.2010.10.002

277. Proverbio AM, Santoni S, Adorni R. ERP Markers of Valence Coding in Emotional Speech Processing. iScience. 2020;23: 100933. doi:10.1016/J.ISCI.2020.100933

278. Remijn GB, Hasuo E, Fujihira H, Morimoto S. An introduction to the measurement of auditory event-related potentials (ERPs). Acoust Sci Technol. 2014;35: 229–242. doi:10.1250/ast.35.229

279. Getz LM, Toscano JC. The time-course of speech perception revealed by temporally-sensitive neural measures. Wiley Interdiscip Rev Cogn Sci. 2021;12. doi:10.1002/WCS.1541

280. Sangal RB, Sangal JM. Topography of Auditory and Visual P300 in Normal Adults. Clin EEG Neurosci. 1996;27: 145–150. doi:10.1177/155005949602700307/ASSET/155005949602700307.FP.PNG_V03

281. Tekok-Kilic A, Shucard JL, Shucard DW. Stimulus modality and Go/NoGo effects on P3 during parallel visual and auditory continuous performance tasks. Psychophysiology. 2001;38: 578–589. doi:10.1017/S0048577201991279

282. Boudewyn MA, Luck SJ, Farrens JL, Kappenman ES. How Many Trials Does It Take to Get a Significant ERP Effect? It Depends. Psychophysiology. 2018;55: e13049. doi:10.1111/PSYP.13049

283. Yano M, Suwazono S, Arao H, Yasunaga D, Oishi H. Inter-participant variabilities and sample sizes in P300 and P600. 2019 [cited 7 Oct 2024]. doi:10.1016/j.ijpsycho.2019.03.010

284. Pontifex MB, Scudder MR, Brown ML, O'Leary KC, Wu CT, Themanson JR, et al. On the number of trials necessary for stabilization of error-related brain activity across the life span. Psychophysiology. 2010;47: 767–773. doi:10.1111/J.1469-8986.2010.00974.X

285. Sharples S, Cobb S, Moody A, Wilson JR. Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. Displays. 2008;29: 58–69. doi:10.1016/j.displa.2007.09.005

286. Dickter CL, Kieffaber PD. EEG methods for the psychological sciences. EEG Methods for the Psychological Sciences. 2014; 1–168. doi:10.4135/9781446270356

287. Weber D, Hertweck S, Alwanni H, Fiederer LDJ, Wang X, Unruh F, et al. A Structured Approach to Test the Signal Quality of Electroencephalography Measurements During Use of Head-Mounted Displays for Virtual Reality Applications. Front Neurosci. 2021;15. doi:10.3389/fnins.2021.733673

288. Motamedi-Fakhr S, Moshrefi-Torbati M, Hill M, Hill CM, White PR. Signal processing techniques applied to human sleep EEG signals-A review. Biomed Signal Process Control. 2014;10: 21–33. doi:10.1016/j.bspc.2013.12.003

289. Hirzle T, Cordts M, Rukzio E, Gugenheimer J, Bulling A. A Critical Assessment of the Use of SSQ as a Measure of General Discomfort in VR Head-Mounted Displays. Proceedings of the 2021 CHI conference on human factors in computing systems. 2021; 1–14. doi:10.1145/3411764

290. Kappenman ES, Luck SJ. The effects of electrode impedance on data quality and statistical significance in ERP recordings. Psychophysiology. 2010;47: 888–904. doi:10.1111/J.1469-8986.2010.01009.X

291. Luck SJ. Basic Principles of ERP Recording. 2nd ed. An Introduction to the Event-Related Potential Technique. 2nd ed. The MIT Press; 2014. pp. 147–183.

292. Moustafa F. Social VR-an Investigation into the Affective Triggers of Virtual Reality. MSc Final Project Report, University College London. 2017.

293. Li G, Chen CP, Pohlmann K, Brewster S, Mcgill M, Pollick F. Exploring Neural Biomarkers in Young Adults Resistant to VR Motion Sickness: A Pilot Study of EEG. 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR). 2023;IEEE: 328–335. doi:10.1109/VR55154.2023.00048

294. Laustsen M, Andersen M, Xue R, Madsen KH, Hanson LG. Tracking of rigid head motion during MRI using an EEG system. Magn Reson Med. 2022;88: 986–1001. doi:10.1002/MRM.29251

295. Lopes P, Tian N, Boulic R. Exploring Blink-Rate Behaviors for Cybersickness Detection in VR. Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VRW 2020. 2020; 795–796. doi:10.1109/VRW50115.2020.00248

296. Ozkan A, Celikcan U. The relationship between cybersickness and eye-activity in response to varying speed, scene complexity and stereoscopic VR parameters. Int J Human-Computer Studies. 2023;176: 103039. doi:10.1016/j.ijhcs.2023.103039

297. Marshev V, Bolloc'h J, Pallamin N, de Bougrenet de la Tocnaye JL, Cochener B, Nourrit V. Impact of virtual reality headset use on eye blinking and lipid layer thickness. J Fr Ophtalmol. 2021;44: 1029–1037. doi:10.1016/J.JFO.2020.09.032

298. Luck SJ. Basics of Fourier Analysis and Filtering. 2nd ed. An Introduction tot he Event-Related Potential Technique. 2nd ed. 2014. pp. 219–248.

299. Li Y, Ma Z, Lu W, Li Y. Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. Physiol Meas. 2006;27: 425. doi:10.1088/0967-3334/27/4/008

300. Hoffmann S, Falkenstein M. The Correction of Eye Blink Artefacts in the EEG: A Comparison of Two Prominent Methods. PLoS One. 2008;3. doi:10.1371/JOURNAL.PONE.0003004

301. Coelli S, Calcagno A, Maria Cassani C, Temporiti F, Reali P, Gatti R, et al. Selecting methods for a modular EEG pre-processing pipeline: An objective comparison. Biomed Signal Process Control. 2023;90: 1746–8094. doi:10.1016/j.bspc.2023.105830

302. Dien J. Best practices for repeated measures ANOVAs of ERP data: Reference, regional channels, and robust ANOVAs. International Journal of Psychophysiology. 2017;111: 42–56. doi:10.1016/J.IJPSYCHO.2016.09.006

303. Brouwer AM, Hogervorst MA, Van Erp JBF, Heffelaar T, Zimmerman PH, Oostenveld R. Estimating workload using EEG spectral power and ERPs in the n-back task. J Neural Eng. 2012;9. doi:10.1088/1741-2560/9/4/045008

304. Peng W. EEG Preprocessing and Denoising. In: Hu L, Zhang Z, editors. EEG Signal Processing and Feature Extraction. 2019. pp. 71–87.

305. VanRullen R. Four common conceptual fallacies in mapping the time course of recognition. Front Psychol. 2011;2: 12780. doi:10.3389/FPSYG.2011.00365/BIBTEX

306. Rousselet GA. Does filtering preclude us from studying ERP time-courses? Front Psychol. 2012;3: 25002. doi:10.3389/FPSYG.2012.00131/BIBTEX

307. de Cheveigné A, Nelken I. Filters: When, Why, and How (Not) to Use Them. Neuron. 2019;102: 280–293. doi:10.1016/J.NEURON.2019.02.039

308. Yael D, Vecht JJ, Bar-Gad I. Filter-Based Phase Shifts Distort Neuronal Timing Information. eNeuro. 2018;5. doi:10.1523/ENEURO.0261-17.2018

309. Acunzo DJ, MacKenzie G, van Rossum MCW. Systematic biases in early ERP and ERF components as a result of high-pass filtering. J Neurosci Methods. 2012;209: 212–218. doi:10.1016/J.JNEUMETH.2012.06.011

310. Tanner D, Morgan-Short K, Luck SJ. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. Psychophysiology. 2015;52: 997. doi:10.1111/PSYP.12437

311. Bigdely-Shamlo N, Mullen T, Kothe C, Su KM, Robbins KA. The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. Front Neuroinform. 2015;9: 1–19. doi:10.3389/FNINF.2015.00016/BIBTEX

312. Leske S, Dalal SS. Reducing power line noise in EEG and MEG data via spectrum interpolation. Neuroimage. 2019;189: 763. doi:10.1016/J.NEUROIMAGE.2019.01.026

313. Widmann A, Schröger E. Filter effects and filter artifacts in the analysis of electrophysiological data. Front Psychol. 2012;3: 233. doi:10.3389/FPSYG.2012.00233/BIBTEX

314. Kober SE, Settgast V, Marlies Brunnhofer, Augsdörfer U, Guilherme Wood. Move your virtual body: differences and similarities in brain activation patterns during hand movements in real world and virtual reality. Virtual Real. 2022;26: 501–511. doi:10.1007/s10055-021-00588-1

315. Jiang X, Bian G-B, Tian Z. Removal of Artifacts from EEG Signals: A Review. 2019 [cited 1 Jan 2024]. doi:10.3390/s19050987

316. Hagemann D, Naumann E. The effects of ocular artifacts on (lateralized) broadband power in the EEG. Article in Clinical Neurophysiology. 2001 [cited 5 Mar 2024]. doi:10.1016/S1388-2457(00)00541-1

317. Luck SJ. Artifact Rejection and Correction. 2nd ed. An introduction to the event-related potential technique. 2nd ed. MIT Press; 2014. pp. 185–217.

318. Makeig S, Jung T-P, Bell AJ, Sejnowski TJ. Independent Component Analysis of Electroencephalographic Data. Adv Neural Inf Process Syst. 1995;8.

319. Onton J, Westerfield M, Townsend J, Makeig S. Imaging human EEG dynamics using independent component analysis. Neurosci Biobehav Rev. 2006;30: 808–822. doi:10.1016/J.NEUBIOREV.2006.06.007

320. Makeig S, Onton J. ERP Features and EEG Dynamics: An ICA Perspective. The Oxford Handbook of Event-Related Potential Components. 2012 [cited 21 Oct 2022]. doi:10.1093/OXFORDHB/9780195374148.013.0035

321. Winkler I, Debener S, Muller KR, Tangermann M. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. Annu Int Conf IEEE Eng Med Biol Soc. 2015;2015: 4101–4105. doi:10.1109/EMBC.2015.7319296

322. Plöchl M, Ossandón JP, König P, Dimigen O, Belyusar D, Einstein A. Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. 2012. doi:10.3389/fnhum.2012.00278

323. Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clinical Neurophysiology. 2000;111: 1745–1758. doi:10.1016/S1388-2457(00)00386-2

324. Perez VB, Vogel EK. What ERPs Can Tell Us about Working Memory. The Oxford Handbook of Event-Related Potential Components. 2011 [cited 13 Mar 2024]. doi:10.1093/OXFORDHB/9780195374148.013.0180

325. Ko LW, Komarov O, Hairston WD, Jung TP, Lin CT. Sustained attention in real classroom settings: An EEG study. Front Hum Neurosci. 2017;11: 269694. doi:10.3389/FNHUM.2017.00388/BIBTEX

326. Gardony AL, Eddy MD, Brunyé TT, Taylor HA. Cognitive strategies in the mental rotation task revealed by EEG spectral power. Brain Cogn. 2017;118: 1–18. doi:10.1016/J.BANDC.2017.07.003

327. Bohbot VD, Copara MS, Gotman J, Ekstrom AD. Low-frequency theta oscillations in the human hippocampus during real-world and virtual navigation. Nature Communications 2017 8:1. 2017;8: 1–7. doi:10.1038/ncomms14415

328. Dickter CL, Kieffaber PD. Frequency-Domain Analysis. EEG Methods for the Psychological Sciences. SAGE Publications Ltd; 2014.

329. Morales S, Bowers ME. Time-frequency analysis methods and their application in developmental EEG data. Dev Cogn Neurosci. 2022;54: 101067. doi:10.1016/J.DCN.2022.101067

330. Zhang Z. Spectral and time-frequency analysis. EEG Signal Processing and Feature Extraction. 2019; 89–116. doi:10.1007/978-981-13-9113-2_6

331. Nunez PL, Srinivasan R. Measures of EEG Dynamic Properties. 2nd ed. In: Nunez PL, Srinivasan R, editors. Electric fields of the brain: The neurophysics of EEG. 2nd ed. Oxford: Oxford Univ. Press; 2006. pp. 353–431.

332. Pfurtscheller G, Lopes Da Silva FH. Event-related EEG/MEG synchronization and desynchronization: Basic principles. Clinical Neurophysiology. 1999;110: 1842–1857. doi:10.1016/S1388-2457(99)00141-8

333. Makeig S. Auditor-Event-Related Dynamics of the EEG Spectrum. Electroencephalogr Clin Neurophysiol. 1993;86: 20. Available: https://sccn.ucsd.edu/~scott/pdf/ERSP93.pdf%0Apapers2://publication/uuid/03A98A40-49C2-4E7A-A91F-B4E0D2B98237

334. Pfurtscheller G. EEG event-related desynchronization (ERD) and synchronization (ERS). Electroencephalogr Clin Neurophysiol. 1997;1: 26. Available: https://www.infona.pl//resource/bwmeta1.element.elsevier-932da8ce-8161-35fe-afd2-3f9f63f14c0c

335. Ameera A, Saidatul A, Ibrahim Z. Analysis of EEG Spectrum Bands Using Power Spectral Density for Pleasure and Displeasure State. IOP Conf Ser Mater Sci Eng. 2019;557: 1–5. doi:10.1088/1757-899X/557/1/012030

336. Corona-González CE, Alonso-Valerdi LM, Ibarra-Zarate DI. Personalized Theta and Beta Binaural Beats for Brain Entrainment: An Electroencephalographic Analysis. Front Psychol. 2021;12. doi:10.3389/FPSYG.2021.764068

337. Wang R, Wang J, Yu H, Wei X, Yang C, Deng B. Power spectral density and coherence analysis of Alzheimer's EEG. Cogn Neurodyn. 2015;9: 291. doi:10.1007/S11571-014-9325-X

338. Holm A, Lukander K, Korpela J, Sallinen M, Müller KMI. Estimating brain load from the EEG. ScientificWorldJournal. 2009;9: 639–651. doi:10.1100/TSW.2009.83

339. Pfurtscheller G, Scherer R, Leeb R, Keinrath C, Neuper C, Lee F, et al. Viewing Moving Objects in Virtual Reality Can Change the Dynamics of Sensorimotor EEG Rhythms. Presence. 2007;16: 111–118.

340. Banaei M, Hatami J, Yazdanfar A, Gramann K. Walking through architectural spaces: The impact of interior forms on human brain dynamics. Front Hum Neurosci. 2017;11: 1–14. doi:10.3389/fnhum.2017.00477

341. Vourvopoulos A, Bermúdezi Badia S. Motor priming in virtual reality can augment motor-imagery training efficacy in restorative brain-computer interaction: A within-subject analysis. J Neuroeng Rehabil. 2016;13: 1–14. doi:10.1186/s12984-016-0173-2

342. Vourvopoulos A, Marin-Pardo O, Neureither M, Saldana D, Jahng E, Liew SL. Multimodal Head-Mounted Virtual-Reality Brain-Computer Interface for Stroke Rehabilitation: A Clinical Case Study with REINVENT. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence

and Lecture Notes in Bioinformatics). 2019;11574 LNCS: 165–179. doi:10.1007/978-3-030-21607-8_13/FIGURES/9

343. Vourvopoulos A, Pardo OM, Lefebvre S, Neureither M, Saldana D, Jahng E, et al. Effects of a brain-computer interface with virtual reality (VR) neurofeedback: A pilot study in chronic stroke patients. Front Hum Neurosci. 2019;13: 460405. doi:10.3389/FNHUM.2019.00210/BIBTEX

344. Dini H, Simonetti A, Bigne E, Bruni LE. EEG theta and N400 responses to congruent versus incongruent brand logos. Scientific Reports 2022 12:1. 2022;12: 1–11. doi:10.1038/s41598-022-08363-1

345. Sanuki F, Nakphu N, Tahara A, Iramina K. The comparison of electroencephalography power and event related potential in success and failure during multitask game. Front Neurorobot. 2022;16: 1044071. doi:10.3389/FNBOT.2022.1044071/BIBTEX

346. Klotzsche F, Gaebler M, Villringer A, Sommer W, Nikulin V, Ohl S. Visual short-term memory-related EEG components in a virtual reality setup. Psychophysiology. 2023;60: e14378. doi:10.1111/PSYP.14378

347. Qadir Z, Chowdhury E, Ghosh L, Konar A. Quantitative Analysis of Cognitive Load Test While Driving in a VR vs Non-VR Environment. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2019;11942 LNCS: 481–489. doi:10.1007/978-3-030-34872-4_53/TABLES/1

348. Albert WS, Reinitz MT, Beusmans JM. The role of attention in spatial learning during simulated route navigation. Environ Plan A. 1999;31: 1459–1472.

349. Brunyé TT, Gardony AL, Holmes A, Taylor HA. Spatial decision dynamics during wayfinding: intersections prompt the decision-making process. Cogn Res Princ Implic. 2018;3: 1–19. doi:10.1186/s41235-018-0098-3

350. Korthauer LE, Nowak NT, Frahmand M, Driscoll I. Cognitive correlates of spatial navigation: Associations between executive functioning and the virtual Morris Water Task. Behavioural Brain Research. 2017;317: 470–478. doi:10.1016/j.bbr.2016.10.007

351. Garden S, Cornoldi C, Logie RH. Visuo-spatial working memory in navigation. Appl Cogn Psychol. 2002;16: 35–50. doi:10.1002/ACP.746

352. Baumann O, Skilleter AJ, Mattingley JB. Short-Term Memory Maintenance of Object Locations during Active Navigation: Which Working Memory Subsystem Is Essential? PLoS One. 2011;6: e19707. doi:10.1371/JOURNAL.PONE.0019707

353. Blacker KJ, Weisberg SM, Newcombe NS, Courtney SM. Keeping Track of Where We Are: Spatial Working Memory in Navigation. Vis cogn. 2017;25: 691. doi:10.1080/13506285.2017.1322652

354. Plechatá A, Sahula V, Fayette D, Fajnerová I. Age-related differences with immersive and non-immersive virtual reality in memory assessment. Front Psychol. 2019;10: 434210. doi:10.3389/FPSYG.2019.01330/BIBTEX

355. Kunishige M, Miyaguchi H, Fukuda H, Iida T, Nami K, Ishizuki C. Spatial navigation ability is associated with the assessment of smoothness of driving during changing lanes in older drivers. J Physiol Anthropol. 2020;39: 1–11. doi:10.1186/S40101-020-00227-9/FIGURES/8

356. Nys M, Hickmann M, Gyselinck V. The role of verbal and visuo-spatial working memory in the encoding of virtual routes by children and adults. Journal of Cognitive Psychology. 2018;30: 710–727. doi:10.1080/20445911.2018.1523175

357. Ekstrom AD, Huffman DJ, Starrett M. Interacting networks of brain regions underlie human spatial navigation: A review and novel synthesis of the literature. J Neurophysiol. 2017;118: 3328–3344. doi:10.1152/JN.00531.2017/ASSET/IMAGES/LARGE/Z9K0121744050004.JPEG

358. Epstein RA, Patai EZ, Julian JB, Spiers HJ. The cognitive map in humans: Spatial navigation and beyond. Nat Neurosci. 2017;20: 1504. doi:10.1038/NN.4656

359. Wiener JM, Carroll D, Moeller S, Bibi I, Ivanova D, Allen P, et al. A novel virtual-reality-based route-learning test suite: Assessing the effects of cognitive aging on navigation. 2019 [cited 9 Mar 2024]. doi:10.3758/s13428-019-01264-8

360. Chersi F, Burgess N. The Cognitive Architecture of Spatial Navigation: Hippocampal and Striatal Contributions. Neuron. 2015;88: 64–77. doi:10.1016/J.NEURON.2015.09.021

361. Coluccia E. Learning from maps: The role of visuo-spatial working memory. Appl Cogn Psychol. 2008;22: 217–233. doi:10.1002/ACP.1357

362. Muffato V, Meneghetti C, De Beni R. Spatial mental representations: the influence of age on route learning from maps and navigation. Psychol Res. 2019;83: 1836–1850. doi:10.1007/s00426-018-1033-4

363. Harris MA, Wolbers T. How age-related strategy switching deficits affect wayfinding in complex environments. Neurobiol Aging. 2014;35: 1095–1102. doi:10.1016/J.NEUROBIOLAGING.2013.10.086

364. Meneghetti C, Zancada-Menéndez C, Sampedro-Piquero P, Lopez L, Martinelli M, Ronconi L, et al. Mental representations derived from navigation: The role of visuo-spatial abilities and working memory. Learn Individ Differ. 2016;49: 314–322. doi:10.1016/J.LINDIF.2016.07.002

365. Gras D, Gyselinck VV, Perrussel M, Orriols E, Piolino P. The role of working memory components and visuospatial abilities in route learning within a virtual environment. Journal of Cognitive Psychology. 2013;25: 38–50. doi:10.1080/20445911.2012.739154

366. Lithfous S, Tromp D, Dufour A, Pebayle T, Goutagny R, Després O. Decreased theta power at encoding and cognitive mapping deficits in elderly individuals during a spatial memory task. Neurobiol Aging. 2015;36: 2821–2829. doi:10.1016/J.NEUROBIOLAGING.2015.07.007

367. Herweg NA, Kahana MJ. Spatial representations in the human brain. Front Hum Neurosci. 2018;12: 366579. doi:10.3389/FNHUM.2018.00297/BIBTEX

368. Cheng B, Wunderlich A, Gramann K, Lin E, Fabrikant SI. The effect of landmark visualization in mobile maps on brain activity during navigation: A

virtual reality study. Front Virtual Real. 2022;3: 981625. doi:10.3389/FRVIR.2022.981625/BIBTEX

369. Kahana MJ, Sekuler R, Caplan JB, Kirschen M, Madsen JR. Human theta oscillations exhibit task dependence during virtual maze navigation. Nature 1999 399:6738. 1999;399: 781–784. doi:10.1038/21645

370. Kober SE, Kurzmann J, Neuper C. Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: An EEG study. International Journal of Psychophysiology. 2012;83: 365–374. doi:10.1016/J.IJPSYCHO.2011.12.003

371. Slobounov SM, Ray W, Johnson B, Slobounov E, Newell KM. Modulation of cortical activity in 2D versus 3D virtual reality environments: An EEG study. International Journal of Psychophysiology. 2015;95: 254–260. doi:10.1016/J.IJPSYCHO.2014.11.003

372. Chrastil ER, Rice C, Goncalves M, Moore KN, Wynn SC, Stern CE, et al. Theta oscillations support active exploration in human spatial navigation. Neuroimage. 2022;262: 119581. doi:10.1016/J.NEUROIMAGE.2022.119581

373. Jeung S, Hilton C, Berg T, Gehrke L, Gramann K. Virtual Reality for Spatial Navigation. Curr Top Behav Neurosci. 2023;65: 103–129. doi:10.1007/7854_2022_403

374. Zisch FE, Newton C, Coutrot A, Murcia M, Motala A, Greaves J, et al. Comparable human spatial memory distortions in physical, desktop virtual and immersive virtual environments. bioRxiv. 2022; 2022.01.11.475791. doi:10.1101/2022.01.11.475791

375. Rudolph B, Musick G, Wiitablake L, Lazar KB, Mobley C, Boyer DM, et al. Investigating the Effects of Display Fidelity of Popular Head-Mounted Displays on Spatial Updating and Learning in Virtual Reality. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2020;12509 LNCS: 666–679. doi:10.1007/978-3-030-64556-4_52/FIGURES/3

376. Nguyen Do T-T, Lin C-T, Gramann K. Human brain dynamics in active spatial navigation. Sci Rep. 2021;11: 13036. doi:10.1038/s41598-021-92246-4

377. Wen D, Yuan J, Li J, Sun Y, Wang X, Shi R, et al. Design and Test of Spatial Cognitive Training and Evaluation System Based on Virtual Reality Head-Mounted Display With EEG Recording. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2023;31: 2705–2714. doi:10.1109/TNSRE.2023.3283328

378. Liu J, Singh AK, Lin CT. Using virtual global landmark to improve incidental spatial learning. Scientific Reports 2022 12:1. 2022;12: 1–14. doi:10.1038/s41598-022-10855-z

379. Zhu B, Cruz-Garza JG, Yang Q, Shoaran M, Kalantari S. Identifying uncertainty states during wayfinding in indoor environments: An EEG classification study. Advanced Engineering Informatics. 2022;54: 101718. doi:10.1016/J.AEI.2022.101718

380. Du YK, Liang M, Mcavan AS, Wilson RC, Ekstrom AD. Frontal-midline theta and posterior alpha oscillations index early processing of spatial representations during active navigation. 2023 [cited 9 Mar 2024]. doi:10.1016/j.cortex.2023.09.005

381. Ruddle RA, Payne SJ, Jones DM. Navigating large-scale virtual environments: What differences occur between helmet-mounted and desk-top displays? Presence: Teleoperators and Virtual Environments. 1999;8: 157–168. doi:10.1162/105474699566143

382. Marraffino MD, Johnson CI, Garibaldi AE. Virtual Reality is Better Than Desktop for Training a Spatial Knowledge Task, but Not for Everyone. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2022;13317 LNCS: 212–223. doi:10.1007/978-3-031-05939-1_14/FIGURES/3

383. Dong W, Qin T, Yang T, Liao H, Liu B, Meng L, et al. Wayfinding Behavior and Spatial Knowledge Acquisition: Are They the Same in Virtual Reality and in Real-World Environments? Ann Am Assoc Geogr. 2022;112: 226–246. doi:10.1080/24694452.2021.1894088

384. Boulanger P, Torres D, Bischof W. MANDALA: A Reconfigurable VR Environment for Studying Spatial Navigation in Humans Using EEG. Eurographics Symposium on Virtual Environments. 2004.

385. Kober SE, Neuper C. Sex differences in human EEG theta oscillations during spatial navigation in virtual reality. International Journal of Psychophysiology. 2011;79: 347–355. doi:10.1016/j.ijpsycho.2010.12.002

386. Marples D, Gledhill D, Carter P. The Effect of Light-ing, Landmarks and Auditory Cues on Human Performance in Navigating a Virtual Maze. 2020 [cited 10 Mar 2024]. doi:10.1145/3384382.3384527

387. McClendon MS. The Complexity and Difficulty of a Maze. Bridges: Mathematical connections in art, music, and science. 2001; 213–222.

388. Meneghetti C, Labate E, Toffalini E, Pazzaglia F. Successful navigation: the influence of task goals and working memory. Psychol Res. 2021;85: 634–648. doi:10.1007/S00426-019-01270-7/FIGURES/5

389. Bartlett ML, Gwinn OS, Thomas NA, Nicholls MER. Cognitive load exacerbates rightward biases during computer maze navigation. 2020 [cited 1 Mar 2024]. doi:10.1016/j.bandc.2020.105547

390. Palombi T, Mandolesi L, Alivernini F, Chirico A, Lucidi F. Application of Real and Virtual Radial Arm Maze Task in Human. Brain Sciences 2022, Vol 12, Page 468. 2022;12: 468. doi:10.3390/BRAINSCI12040468

391. Olton DS, Collison C, Werz MA. Spatial memory and radial arm maze performance of rats. Learn Motiv. 1977;8: 289–314. doi:10.1016/0023-9690(77)90054-6

392. Suarez JA, Solano JL, Barrios KP, Ortega LA. Nicotine increases behavioral variability on radial arm maze extinction. A preliminary study. Learn Motiv. 2021;74: 101721. doi:10.1016/J.LMOT.2021.101721

393. Erkan İ. Examining wayfinding behaviours in architectural spaces using brain imaging with electroencephalography (EEG). Archit Sci Rev. 2018;61: 410–428. doi:10.1080/00038628.2018.1523129

394. Sharma G, Kaushal Y, Chandra S, Singh V, Mittal AP, Dutt V. Influence of landmarks on wayfinding and brain connectivity in immersive virtual reality environment. Front Psychol. 2017;8: 1–12. doi:10.3389/fpsyg.2017.01220

395. Suma E, Finkelstein S, Reid M, Babu S, Ulinski A, Hodges LF. Evaluation of the cognitive effects of travel technique in complex real and virtual environments. IEEE Trans Vis Comput Graph. 2010;16: 690–702. doi:10.1109/TVCG.2009.93

396. Basu A, Johnsen K. Navigating a maze differently-a user study. 2018.

397. Cao A, Wang L, Liu Y, Popescu V. Feature Guided Path Redirection for VR Navigation. 2020; 137–145. doi:10.1109/VR46266.2020.00032

398. Sousa Santos B, Dias P, Pimentel A, Baggerman JW, Ferreira C, Silva S, et al. Head-mounted display versus desktop for 3D navigation in virtual reality: A user study. Multimed Tools Appl. 2009;41: 161–181. doi:10.1007/S11042-008-0223-2/FIGURES/9

399. Gazova I, Laczó J, Rubinova E, Mokrisova I, Hyncicova E, Andel R, et al. Spatial navigation in young versus older adults. Front Aging Neurosci. 2013;5. doi:10.3389/FNAGI.2013.00094

400. Van Der Ham IJM, Claessen MHG. How age relates to spatial navigation performance: Functional and methodological considerations. 2020 [cited 9 Oct 2024]. doi:10.1016/j.arr.2020.101020

401. Berka C, Levendowski DJ, Lumicao MN, Yau A, Davis G, Zivkovic VT, et al. EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviat Space Environ Med. 2007;5: B231–B244.

402. Saredakis D, Szpak A, Birckhead B, Keage HAD, Rizzo A, Loetscher T. Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis. Front Hum Neurosci. 2020;14: 512264. doi:10.3389/FNHUM.2020.00096/BIBTEX

403. Nasiri M, Porter J, Kohm K, Robb A. Changes in Navigation over Time: A Comparison of Teleportation and Joystick-Based Locomotion. ACM Trans Appl Percept. 2023;20. doi:10.1145/3613902

404. Wais PE, Arioli M, Anguera-Singla R, Gazzaley A. Virtual reality video game improves high-fidelity memory in older adults. Scientific Reports 2021 11:1. 2021;11: 1–14. doi:10.1038/s41598-021-82109-3

405. Jensen O, Tesche CD. Frontal theta activity in humans increases with memory load in a working memory task. European Journal of Neuroscience. 2002;15: 1395–1399. doi:10.1046/J.1460-9568.2002.01975.X

406. Stipacek A, Grabner RH, Neuper C, Fink A, Neubauer AC. Sensitivity of human EEG alpha band desynchronization to different working memory components and increasing levels of memory load. Neurosci Lett. 2003;353: 193–196. doi:10.1016/J.NEULET.2003.09.044

407. Maurer U, Brem S, Liechti M, Maurizio S, Michels L, Brandeis D. Frontal Midline Theta Reflects Individual Task Performance in a Working Memory Task. Brain Topogr. 2015;28: 127–134. doi:10.1007/S10548-014-0361-Y/FIGURES/3

408. Cave AE, Barry RJ. Sex differences in resting EEG in healthy young adults. International Journal of Psychophysiology. 2021;161: 35–43. doi:10.1016/J.IJPSYCHO.2021.01.008

409. Caplan JB, Madsen JR, Raghavachari S, Kahana MJ. Distinct patterns of brain oscillations underlie two basic parameters of human maze learning. J Neurophysiol. 2001;86: 368–380. doi:10.1152/JN.2001.86.1.368

410. Jang KM, Kwon M, Nam SG, Kim DM, Lim HK. Estimating objective (EEG) and subjective (SSQ) cybersickness in people with susceptibility to motion sickness. Appl Ergon. 2022;102: 103731. doi:10.1016/J.APERGO.2022.103731

411. Naqvi SAA, Badruddin N, Jatoi MA, Malik AS, Hazabbah W, Abdullah B. EEG based time and frequency dynamics analysis of visually induced motion sickness (VIMS). Australas Phys Eng Sci Med. 2015;38: 721–729. doi:10.1007/S13246-015-0379-9/FIGURES/5

412. Krokos E, Varshney A. Quantifying VR cybersickness using EEG. Virtual Real. 2022;26: 77–89. doi:10.1007/S10055-021-00517-2/FIGURES/12

413. Pavlov YG, Kasanov D, Kosachenko AI, Kotyusov AI, Busch NA. Pupillometry and electroencephalography in the digit span task. Scientific Data 2022 9:1. 2022;9: 1–6. doi:10.1038/s41597-022-01414-2

414. Hsieh TJT, Kuo YH, Niu CK. Utilizing HMD VR to improve the spatial learning and wayfinding effects in the virtual maze. Communications in Computer and Information Science. 2018;852: 38–42. doi:10.1007/978-3-319-92285-0_6/FIGURES/5

415. Taheri SM, Matsushita K, Sasaki M. Development of a Driving Simulator with Analyzing Driver's Characteristics Based on a Virtual Reality Head Mounted Display. J Transp Technol. 2017;07: 351–366. doi:10.4236/JTTS.2017.73023

416. Affanni A, Najafi TA, Guerci S. Development of an EEG Headband for Stress Measurement on Driving Simulators. Sensors 2022, Vol 22, Page 1785. 2022;22: 1785. doi:10.3390/S22051785

417. Berkman Mİ, Çatak G, Eremektar MÇ. Comparison of VR and Desktop Game User Experience in a Puzzle Game: "Keep Talking and Nobody Explodes." AJIT-e: Academic Journal of Information Technology. 2020;11: 180–204. doi:10.5824/AJITE.2020.03.008.X

418. GomezRomero-Borquez J, Del Puerto-Flores JA, Del-Valle-Soto C. Mapping EEG Alpha Activity: Assessing Concentration Levels during Player Experience in Virtual Reality Video Games. Future Internet 2023, Vol 15, Page 264. 2023;15: 264. doi:10.3390/FI15080264

419. Carroll M, Osborne E, Yildirim C. Effects of VR gaming and game genre on player experience. 2019 IEEE Games, Entertainment, Media Conference, GEM 2019. 2019. doi:10.1109/GEM.2019.8811554

420. Commins S, Duffin J, Chaves K, Leahy D, Corcoran K, Caffrey M, et al. NavWell: A simplified virtual-reality platform for spatial navigation and memory experiments. Behav Res Methods. 2020;52: 1189–1207. doi:10.3758/S13428-019-01310-5/FIGURES/7

421. Brookes J, Warburton M, Alghadier M, Mon-Williams M, Mushtaq F. Studying human behavior with virtual reality: The Unity Experiment Framework. Behav Res Methods. 2020;52: 455–463. doi:10.3758/S13428-019-01242-0

422. Wang C. Comprehensively Summarizing What Distracts Students from Online Learning: A Literature Review. Hum Behav Emerg Technol. 2022;2022. doi:10.1155/2022/1483531

423. Borrego A, Latorre J, Alcañiz M, Llorens R. Comparison of Oculus Rift and HTC Vive: Feasibility for Virtual Reality-Based Exploration, Navigation, Exergaming, and Rehabilitation. Games Health J. 2018;7: 151–156. doi:10.1089/G4H.2017.0114/ASSET/IMAGES/LARGE/FIGURE3.JPEG

424. Chang E, Billinghurst M, Yoo B. Brain activity during cybersickness: a scoping review. Virtual Real. 2023;27: 2073–2097. doi:10.1007/S10055-023-00795-Y/FIGURES/9

425. Caserman P, Garcia-Agundez A, Alvar ·, Zerban G, Göbel S. Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. 2021;25: 1153–1170. doi:10.1007/s10055-021-00513-6

426. Langbehn E, Lubos P, Steinicke F. Evaluation of locomotion techniques for room-scale VR: Joystick, teleportation, and redirected walking. ACM International Conference Proceeding Series. 2018 [cited 15 Mar 2024]. doi:10.1145/3234253.3234291

427. Clifton J, Palmisano S. Effects of steering locomotion and teleporting on cybersickness and presence in HMD-based virtual reality. Virtual Reality 2019 24:3. 2019;24: 453–468. doi:10.1007/S10055-019-00407-8

428. Kim A, Lee JE, Lee KM. Exploring the Relative Effects of Body Position and Locomotion Method on Presence and Cybersickness when Navigating a Virtual Environment. ACM Trans Appl Percept. 2023;21. doi:10.1145/3627706

429. Miyakoshi M, Gehrke L, Gramann K, Makeig S, Iversen J. The AudioMaze: An EEG and motion capture study of human spatial navigation in sparse augmented reality. European Journal of Neuroscience. 2021;54: 8283–8307. doi:10.1111/EJN.15131

430. Lin M-H, Liran O, Bauer N, Baker TE. Power dynamics of theta oscillations during goal-directed navigation in freely moving humans: A mobile EEG-virtual reality T-maze study. bioRxiv. 2021; 2021.10.05.463245. doi:10.1101/2021.10.05.463245

431. Hooks K, Ferguson W, Morillo P, Cruz-Neira C. Evaluating the user experience of omnidirectional VR walking simulators. Entertain Comput. 2020;34: 100352. doi:10.1016/J.ENTCOM.2020.100352

432. Kourtesis P, Papadopoulou A, Roussos P. Cybersickness in Virtual Reality: The Role of Individual Differences, Its Effects on Cognitive Functions and Motor Skills, and Intensity Differences during and after Immersion. Virtual Worlds 2024, Vol 3, Pages 62-93. 2024;3: 62–93. doi:10.3390/VIRTUALWORLDS3010004

433. Apicella A, Barbato S, Chacón LAB, D'Errico G, De Paolis LT, Maffei L, et al. Electroencephalography correlates of fear of heights in a virtual reality environment. Acta IMEKO. 2023;12: 1–7. doi:10.21014/ACTAIMEKO.V12I2.1457

434. Aizenman AM, Koulieris GA, Gibaldi A, Sehgal V, Levi DM, Banks MS, et al. The Statistics of Eye Movements and Binocular Disparities during VR Gaming: Implications for Headset Design. ACM Trans Graph. 2023;42. doi:10.1145/3549529

435. Kardong-Edgren S, Farra SL, Alinier G, Young HM. A Call to Unify Definitions of Virtual Reality. Clin Simul Nurs. 2019;31: 28–34. doi:10.1016/j.ecns.2019.02.006

436. Oberauer K, Jones T, Lewandowsky S. The Hebb repetition effect in simple and complex memory span. Mem Cognit. 2015;43: 852–865. doi:10.3758/S13421-015-0512-8/FIGURES/6

Appendices

Appendix 1) Participant Information Sheet

Participant Information Sheet

Project title	Non-invasive skin surface monitoring of brain and heart activity using integrated mobile systems and virtual reality/augmented reality.
Principal investigator	Name: Dr Aziz Asghar Email address: aziz.asghar@hyms.ac.uk Contact telephone number: 01482 464150
Student investigators	Name: Mr Adam Durnin and Mr Matthew Barras Email addresses: hyad28@hyms.ac.uk, m.r.barras@2014.hull.ac.uk Contact telephone number: 01482 462109 or 01482 462079

What is the purpose of this project?
The purpose of this project is to conveniently and safely record the electrical activity of the brain (brain waves or electroencephalogram, EEG), heart (electrocardiogram, ECG), eyes (electrooculogram, EOG) or muscles (electromyogram, EMG) by placing sensors called electrodes onto the skin surface and using our developed mobile monitoring equipment. We would like to use the mobile monitor systems in healthy participants who are lying down, sitting, standing and walking and compare the results with standard monitoring equipment. Moreover, we may non-invasively measure your skin conductivity (galvanic skin response), temperature and humidity, your pulse oximetry, and use eye-tracking and positional tracking (3D acceleration). We are also interested in studying how the body responds to hearing sounds and seeing images such as still or moving pictures. You may be asked to listen to various sounds and/or look at images during different types of movement (lying down, sitting and walking). Sounds will be presented via speaker or headphones. The level of sound volume will be adjusted so that it is comfortable for you. You may be asked to view visually presented information on a screen such as on a smart phone, PC monitor or head-mounted virtual reality or augmented reality headsets. You may also be asked to complete various questionnaires. The investigators will explain in advance what the study will involve. Please do ask the investigators if you have any questions or concerns.

Why have I been chosen?
We are looking for 18-55 year old healthy participants. You have been sent this information because we believe you might fit these requirements.

What happens if I volunteer to take part in this project?
First, it is up to you to decide whether or not to take part. If you decide to take part you will be given this Participant Information Sheet to keep and asked to complete the attached Participant Consent Form. You should give the Participant Consent Form to the investigator at the earliest opportunity. You will also have the opportunity to ask any questions you may have about the project. If you decide to take part you are still free to withdraw at any time and without needing to give a reason. **IMPORTANT** THIS STUDY WILL NOT CLINICALLY EVALUATE ANY INDIVIDUAL PARTICIPANT'S BRAIN, HEART, EYE OR MUSCLE FUNCTION. THE STUDY INVESTIGATORS ARE NOT QUALIFIED TO CLINICALLY INTERPRET THE SIGNALS RECORDED BY THE MONITORS. SHOULD THE INVESTIGATORS NOTICE A FINDING WHICH MAY BE OF CONCERN THE PARTICIPANT'S UK GENERAL PRACTITIONER WILL BE INFORMED OF THE FINDING. IF YOU PREFER US NOT TO CONTACT YOUR UK GENERAL PRACTITIONER, OR YOU ARE NOT REGISTERED WITH A UK GENERAL PRACTITIONER, THEN YOU WILL NOT BE ABLE TO PROCEED IN TAKING PART IN THIS STUDY. NO DIAGNOSTIC OR CLINICIAL ADVICE WILL BE OFFERED BY THE INVESTIGATORS.

What will I have to do?
You will be asked to attend a testing session held in our laboratory at the Loxley Building, University of Hull. Please bring and wear comfortable clothing, and do not use any product in your hair in the morning prior to the experiment. On arrival you will be met by the investigator who will brief you on the testing procedures and answer any questions or concerns that you might have. After signing a Participant Consent Form, the investigator will ask you to complete a questionnaire requesting some information on your present state of health. You will be asked to perform a series of tasks including lying down, sitting, standing and walking while wearing the monitoring equipment. For brain activity monitoring an EEG cap and/or cup electrodes will be placed onto the head, and for the heart ECG monitoring sticky pads are placed on the chest. You may be asked to listen to sounds, look at still or moving images, or respond to tasks presented on a screen including in virtual reality and augmented reality environments. We may also attach a pulse oximetry monitor onto one finger (monitors blood oxygen levels), monitor skin surface temperature, humidity and skin conduction and your level of physical activity may be recorded via gyroscopes/accelerometers. After you have completed the tests, the investigator will give you a debrief sheet explaining the nature of the research, how you can find out about the results, and how you can withdraw your data if you wish. Water will be available for you to drink, although you

will not be able to drink during the time you are performing the testing. It is estimated that the total time to complete this study will be approximately 90 minutes. Shower facilities are available if required.

Will I receive any financial reward or travel expenses for taking part?
No.

Are there any other benefits of taking part?
No.

Will participation involve any physical discomfort or harm?
Sticky pads or electrode gel are routinely used in clinical and research environments and are generally very well tolerated by participants. There is a possible risk of skin reactions to the adhesive in the sticky pads or from the electrode gel. In the unlikely event of this occurring we can try to use sticky pads/gel from a different supplier. There is a possibility that that you may experience nausea/motion sickness when viewing images presented to you on a screen. These symptoms can particularly occur if you are wearing virtual reality or augmented reality headsets. Should you experience nausea/motion sickness please let the researcher know and your participation in the study will be stopped. The device/system we intend to use has been locally independently safety assessed by a colleague in the School of Engineering, University of Hull.

Will I have to provide any bodily samples (e.g. blood or saliva)?
No.

Will participation involve any embarrassment or other psychological stress?
For the ECG electrode placement access to the chest area will be required while this is achieved.

What will happen once I have completed all that is asked of me?
Your data will be anonymised and added to the group results in readiness for a research publication in a peer reviewed journal.

How will my taking part in this project be kept confidential?
You will be allocated an anonymous participant code that will always be used to identify any data that you provide. Your name or other personal details will not be associated with your data. Your consent form and personal details will be stored separately from your data. All paper records will be stored in a locked filing cabinet, accessible only to the research team, and all electronic information will be stored on a password-protected computer and password-protected USB memory stick and encrypted SD card. All

information and data gathered during this research will be stored in line with the 1998 Data Protection Act and will be destroyed 10 years following the conclusion of the study. During that time the data may be used by members of the research team only for purposes appropriate to the research question, but at no point will your personal information or data be revealed.

How will my data be used?

Ultimately your data will be included within the group results in readiness for a research publication in a peer reviewed journal, having been anonymised first.

Who has reviewed this study?

This project has undergone ethical scrutiny and all procedures have been risk assessed and approved by the Hull York Medical School Ethics Committee.

What if I am unhappy during my participation in the project?

You are free to withdraw from the project at any time. During the study itself, if you decide that you do not wish to take any further part then please inform the researcher and they will facilitate your withdrawal. You do not have to give a reason for your withdrawal. Any personal information or data that you have provided (both paper and electronic) will be destroyed or deleted as soon as possible after your withdrawal. After you have completed the research you can still withdraw your personal information and data by contacting the researcher. If you are concerned that regulations are being infringed, or that your interests are otherwise being ignored, neglected or denied, you should inform the Principal Investigator, Dr Aziz Asghar, Email: aziz.asghar@hyms.ac.uk, Telephone: 01482 464150.

How do I take part?

Contact the investigator using the contact details given below. He or she will answer any queries and explain how you can get involved.

Name: Mr Adam Durnin and Mr Matthew Barras

Email addresses: hyad28@hyms.ac.uk, m.r.barras@2014.hull.ac.uk

Contact telephone number: 01482 462109 or 01482 462079

Appendix 2) Participant Consent Form

Participant Consent Form

Project title	Non-invasive skin surface monitoring of brain and heart activity using integrated mobile systems and virtual reality/augmented reality.
Principal investigator	Name: Dr Aziz Asghar Email address: aziz.asghar@hyms.ac.uk Contact telephone number: 01482 464150
Student investigators	Name: Mr Adam Durnin and Mr Matthew Barras Email addresses: hyad28@hyms.ac.uk, m.r.barras@2014.hull.ac.uk Contact telephone number: 01482 462109 or 01482 462079

Please Initial

1. I confirm that I have read and understood all the information provided in the Participant Information Sheet relating to the above project and I have had the opportunity to ask questions. ☐

2. I understand this project is designed to further scientific knowledge and that all procedures have been approved by the Hull York Medical School Ethics Committee. Any questions I have about my participation in this project have been answered to my satisfaction. ☐

3. I fully understand my participation is voluntary and that I am free to withdraw from this project at any time and at any stage, without giving any reason. ☐

4. I understand that the investigators are not qualified to clinically interpret the signals recorded by the monitors. I give my consent that my UK General Practitioner be contacted should the investigators become aware of a finding which may be a cause for concern. ☐

5. I understand that no diagnostic or clinical advice will be offered by the investigators. ☐

6. I have read and fully understand this Participant Consent Form. I agree to take part in this project. ☐

Name of participant **Date** **Signature**

....

Person taking consent **Date** **Signature**

....

UK General Practitioner Address (*If you prefer us not to contact your UK General Practitioner, or you are not registered with a UK General Practitioner, then you will not be able to proceed with the study.*)

.....

Appendix 3) Participant Debrief Form

Participant Debrief Form

Project title	Non-invasive skin surface monitoring of brain and heart activity using integrated mobile systems and virtual reality/augmented reality.
Principal investigator	Name: Dr Aziz Asghar Email address: aziz.asghar@hyms.ac.uk Contact telephone number: 01482 464150
Student investigators	Name: Mr Adam Durnin and Mr Matthew Barras Email addresses: hyad28@hyms.ac.uk, m.r.barras@2014.hull.ac.uk Contact telephone number: 01482 462109 or 01482 462079

What was the purpose of the project?

Our general aim is to determine whether quality datasets can be captured using ambulatory recording systems when presenting cognitive tasks on various screens including within virtual reality (VR) and augmented reality (AR).

How will I find out about the results?

The results will be grouped with those of other study participants and the data published in a peer reviewed journal.

Will I receive any individual feedback?

Feedback to individual participants will not be available as the data will be considered in a group context.

What will happen to the information I have provided?

Your data will be stored safely, will remain confidential, and will be destroyed after 10 years. The data may be shared amongst different researchers but only for the purposes of research, but in all cases we will ensure appropriate confidentiality.

How will the results be disseminated?

The results may be published in a scientific journal or presented at a conference. The data will be generalised, so that your individual data and personal information will not be identifiable.

Have I been deceived in any way during the project?
No.

If I change my mind and wish to withdraw the information I have provided, how do I do this?
Please send an email to aziz.asghar@hyms.ac.uk or a letter (Dr Aziz Asghar) stating that you would like to withdraw your personal information and data from the study. You do not need to give a reason for withdrawing.

Appendix 4) Health Questionnaire

Health Questionnaire

The information in this document will be treated as strictly confidential

Name: .....

Date of Birth: ..... Age: ..... Sex: .....

Blood pressure: …… Resting Heart Rate: …….

Height (cm): ……..... Weight (Kg): ……...

Occupation: .....

Please answer the following questions by putting a circle round the appropriate response or filling in the blank.

1. Do you have, or have you ever had a:

Neurological condition e.g. epilepsy	Yes/No
Heart complaint/ heart disease	Yes/No
Heart pacemaker	Yes/No
Neuro-stimulator	Yes/No

If you answered **Yes**, please do not participate further in this study.

2. How would you describe your present level of **exercise** activity?
   Sedentary / Moderately active / Active / Highly active

3. Please outline a typical weeks exercise activity

…….....
…….....
…….....

4. How would you describe your present level of **lifestyle** activity?
   Sedentary / Moderately active / Active / Highly active

5. How would you describe your present level of fitness?
   Unfit / Moderately fit / Trained / Highly trained

6. Smoking Habits

Are you currently a smoker?	Yes / No
How many cigarettes do you smoke	…….. per day
Are you a previous smoker?	Yes / No
How long is it since you stopped?	..... years
How many did you smoke?	..... per day

7. Do you drink alcohol? Yes / No

If you answered **Yes** and you are male do you drink more than 28 units a week?

Yes / No

If you answered **Yes** and you are female do you drink more than 21 units a week?

Yes / No

8. Have you had to consult your doctor within the last six months? Yes / No

If you answered **Yes**, Have you been advised **not** to exercise?

Yes / No

9. Are you presently taking any form of medication? Yes / No

If you answered **Yes**, Have you been advised **not** to exercise?

Yes / No

10. Do you currently have any form of muscle or joint injury? Yes / No

If you answered **Yes**, please give details……..
……
…….....

11. Are you suffering from any known serious infection? Yes / No

12. As far as you are aware, is there anything that might prevent you from successfully completing the tests that have been outlined to you? Yes / No.

If you answered **Yes**, please give details……..
……
…….....……

PLEASE SIGN AND DATE

Participant's Signature: ……
Date……

Investigator's Signature:……..
Date……

Appendix 5) Arithmetic Experiment Preliminary Questionnaire

Preliminary Questionnaire

1. Do you have any experience using virtual reality headsets (e.g. Oculus Rift, PSVR, Google Cardboard)? If so, please list the headsets you have used:

   .....

   .....

   .....

   .....

2. If you have answered yes to question 1, please circle the amount of time you have used VR for:

   Never < 1 Hour 1 Hour to 10 Hours 10 to 24 Hours > 24 Hours

3. Please circle your preferences in the use of hands in the following activities or objects:

Writing	Always Right	Usually Right	Both Equally	Usually Left	Always Left
Throwing	Always Right	Usually Right	Both Equally	Usually Left	Always Left
Toothbrush	Always Right	Usually Right	Both Equally	Usually Left	Always Left
Spoon	Always Right	Usually Right	Both Equally	Usually Left	Always Left

Appendix 6) Paper Version of the Simulator Sickness Questionnaire

Simulator Sickness Questionnaire

Instructions: Circle how much of each symptom is affecting you right now

General Discomfort	None	Slight	Moderate	Severe
Fatigue	None	Slight	Moderate	Severe
Headache	None	Slight	Moderate	Severe
Eye Strain	None	Slight	Moderate	Severe
Difficulty Focusing	None	Slight	Moderate	Severe
Salivation Increasing	None	Slight	Moderate	Severe
Sweating	None	Slight	Moderate	Severe
Nausea	None	Slight	Moderate	Severe
Difficulty Concentrating	None	Slight	Moderate	Severe
Fullness of the Head/ Head Feels Heavy	None	Slight	Moderate	Severe
Blurred Vision	None	Slight	Moderate	Severe
Dizziness with Eyes Open	None	Slight	Moderate	Severe
Dizziness with Eyes Closed	None	Slight	Moderate	Severe
Vertigo	None	Slight	Moderate	Severe
Stomach Awareness/ Stomach Heaviness	None	Slight	Moderate	Severe
Burping	None	Slight	Moderate	Severe

Appendix 7) Preliminary Questionnaire for the Spatial Navigation Experiment

Preliminary Questionnaire

1. Do you have any experience using virtual reality headsets (e.g. Oculus Rift, PSVR, Google Cardboard)? If so, please list the headsets you have used:

   .....

   .....

2. If you have answered yes to question 1, please circle the amount of time you have used VR for:

   Never < 1 Hour 1 Hour to 10 Hours 10 to 24 Hours > 24 Hours

3. Please circle your preferences in the use of hands in the following activities or objects:

Writing	Always Right	Usually Right	Both Equally	Usually Left	Always Left
Throwing	Always Right	Usually Right	Both Equally	Usually Left	Always Left
Toothbrush	Always Right	Usually Right	Both Equally	Usually Left	Always Left
Spoon	Always Right	Usually Right	Both Equally	Usually Left	Always Left

4. Other than VR, do you play video games? If so, how many hours per week on average do you play:

   Never < 1 Hour 1 Hour to 7 Hours 7 to 14 Hours > 14 Hours

5. Please circle the devices you use to play games on

   Console (Xbox, PlayStation, Switch) PC Mobile Other (please write below)

6. Of this, how much of this is played using a gamepad/controller? For example, an Xbox controller.

Never 1-20% 20-40% 40-60% 60-80% 80-100%

7. Please list the genre of games you mostly play (first person shooter, racing, etc.:)

.....

.....