University of Plymouth

PEARL **https://pearl.plymouth.ac.uk**

04 University of Plymouth Research Theses 01 Research Theses Main Collection

2023

Utilising Event Cognition Principles to Explore the Effects of Location Change on Memory Within Immersive Virtual Environments

Watson, Paul

https://pearl.plymouth.ac.uk/handle/10026.1/20610

http://dx.doi.org/10.24382/2664
University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior consent.

UNIVERSITY OF
PLYMOUTH

UTILISING EVENT COGNITION PRINCIPLES TO EXPLORE THE EFFECTS OF LOCATION CHANGE ON MEMORY WITHIN IMMERSIVE VIRTUAL ENVIRONMENTS

by

PAUL WILLIAM LINDSAY WATSON

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Engineering, Computing and Mathematics

MARCH 2023

Acknowledgements

I am exceptionally grateful to those that have supported me through the work presented in this thesis. I hope all those that have given their time for feedback, guidance and cathartic discussion can see their influence on the work presented. These interactions have shaped the content of this work and allowed me to glimpse the various reasons why research is important to so many.

Firstly, I would like to thank the School of Engineering, Computing, and Mathematics at the University of Plymouth for sponsoring my place on the PhD program. Without this support, the work presented could not be developed.

The biggest thanks to Dr Swen Guadl for the many hours of guidance and interest you have shown in this work. Your mentorship, academic support, and pastoral care have been remarkable. Dr Dan Livingstone and Luke Angel. Thank you for all the passion and reflection at the beginning of this journey. I will always be grateful.

A big thank you to, Professor Sabine Pahl, Dr Marius Golubickis, and Professor Jon May for your detailed approach to statistics and a keen interest in what drives human behaviour. Not only was the psychology perspective fascinating and critical in developing this thesis, but your approach to methodological rigour and analysis was inspirational. I hope you will see these lessons well applied in any future work.

To my partner, Magda, thank you for all the love and support during this thesis. I cannot thank you enough for giving me the time I needed to develop this work and the belief that no challenge was insurmountable. I guess that is a nice way of saying thank you for putting up with me during a very hectic time. My children Evie and Noah, you both came into our lives during this thesis. Your smiles and insight into the world are beautiful. Keep on being interested in the world and talking to your parents!

Importantly I thank my friends and colleagues who joined me on this journey. Dr Marius Varga, your council has always been astute and your humour a bridge to take on the next challenge. I am looking forward to our many chats in the future. To Laura and Sophie, thank you for keeping it real and giving my family the bests of times through what has been the busiest of times. James Jarvis, your support was invaluable. With you, technology works.

Author's declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award. Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

Relevant scientific seminars and conferences were regularly attended at which work was often presented:

Publications:

Watson, P. and Livingstone, D., 2018. Using mixed reality displays for observational learning of motor skills: A design research approach enhancing memory recall and usability. Research in Learning Technology, 26.

Watson, P. and Gaudl, S., 2021. Walking through virtual doors: A study on the effects of virtual location changes on memory. International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments (2021) J. Orlosky, D. Reiners, and B. Weyers (Editors)

Word count for the main body of this thesis: **36,930**

Signed: ____________________

Date: 20/03/2023

Abstract

Paul William Lindsay Watson

Utilising Event Cognition Principles to Explore the Effects of Location Change on Memory Within Immersive Virtual Environments

The thesis covers a series of studies that explore and investigate the application of event cognition theory to the design of virtual reality (VR) spaces for improved memory recall. The thesis starts with a usability study of a recovery position VR training app developed within the University of Plymouth (Watson & Livingstone, 2018; Watson & Gaudl, 2021). This work compares the recall and satisfaction of using this application across mobile VR, desktop displays, and pre-recorded video. Based on the first prototype the question then posed is:

"Can we better design immersive virtual training spaces that support cognitive processes that might aid learning."

Although enhancing memory does not necessarily equate to more learning, recall of information is an important step in many approaches to education and training. Previous real world event cognition work has observed an increase in recall when separating information between rooms, compared to having the same information delivered within a single room (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016). This enhancement comes without any specified strategy from the subjects. Segmenting information between VR rooms might also observe similar memory benefits. VR allows complete control over the visual and audio feed to the user, the narrative of events, and affords spatial understanding similar to the real world.

Across a further three studies, the thesis explores if recall can be improved by separating information between two immersive virtual rooms: In study 1, using a repeated measures design, subjects are exposed to two word lists within one VR room or between VR two rooms (Watson & Gaudl, 2021). In study 2, using a repeated measures design, subjects are exposed to two word lists within one room or between two rooms. However, both of these conditions are performed in both the real world and the immersive VR world.In study 3, using a repeated measures design, subjects are exposed to two word lists within one room or between two rooms. However, this study also controls for variation of the local and global scene upon word lists delivery. For example whether or not the word lists are delivered with the same voice and aesthetic, or variation within these characteristics. Based on our incremental study design, this work observes no significant benefit for segmenting wordlists between immersive virtual rooms for the population sample presented. To support these studies, a virtual reality tool is developed and refined to teach interaction and navigation paradigms, experimental procedures, and facilitate a realistic cognitive simulation of room environments through believable and distinct aesthetics. Post experiment surveys suggest that the virtual reality tool if fit for purpose as a tutorial tool and to deliver the virtual experimental conditions.

Contents

Acknowledgements **1**

Author's declaration **3**

Abstract **5**

Table of Contents **7**

List of Figures **11**

List of Tables **15**

Introduction **17**

Motivation 17

Research Questions 18

Overview of experiments 19

Overview of results 20

Contributions 20

1 Background **23**

1.1 Chapter Summary 23

1.2 Can Location Change Mediate Recall? 24

1.3 Serial Position Curves of List Recall 29

1.4 Virtual Reality Affords Spatial Cognition 31

1.5 Challenges When Using VR For Experimental Trials 35

1.6 Summary 38

2 Preliminary Study **39**

2.1 Chapter Summary 39

2.2 Motivation 40

2.3 Introduction 40

2.4 Design of the Recovery Position Application 43

2.4.1 Hardware Considerations of Mobile Virtual Reality 44

2.4.2 Delivery of Observational Content 46
2.5 Methodology 49
2.5.1 Procedure 49
2.5.2 Data Analysis 52
2.6 Results 54
2.6.1 Participants 54
2.6.2 Recall 55
2.6.3 Usability 56
2.6.4 Autonomy 57
2.6.5 Interviews 58
2.7 Discussion 60
2.8 Conclusion 64
2.9 Future Directions 65
2.10 Limitations 66

3 Experiment 01 67

3.1 Chapter Summary 67
3.2 Introduction 67
3.3 Background 68
3.4 Methodology 72
3.5 Results 75
3.6 Discussion 78
3.7 Limitations and Future Work 82
3.8 Conclusions 83

4 Experiment 02 85

4.1 Chapter Summary 85
4.2 Method 86
4.2.1 Participants 86
4.2.2 Grouping 86
4.2.3 Interaction and Delivery 86
4.2.4 Procedure 87
4.2.5 Learning the Virtual Controls 88
4.2.6 Hardware and Software 89
4.3 Results 90
4.4 Discussion 97
4.5 Conclusion 99
4.6 Limitations 100

5 **Experiment 03** **101**
5.1 Chapter Summary 101
5.2 Method 102
5.2.1 Materials 104
5.2.2 Procedure 112
5.2.3 Timings 114
5.3 Pilot 114
5.4 Results 115
5.4.1 Participants 115
5.4.2 Recall 117
5.4.3 VR System 120
5.5 Discussion 125
5.6 Conclusion 128
5.7 Limitations 128

Conclusion **129**
Can we utilise the segmentation of information between locations as an environment design approach to aid memory recall for VR training? 131
What are the challenges for replicating real-world experiments within VR? . . . 137
Future Work 140

Appendices **145**

A Supplementary Material **146**

Bibliography **149**

List of Figures

1.1 Example serial position curve for descriptive purposes only, and not based on a specific data set. This illustrates the primacy and recency effects of word list recall. The beginning and end of a delivered word list will be recalled more often compared to the middle items when the list is unstructured, and the recall test occurs straight after delivery. The primacy effect describes the increase in recall at the beginning of the list and the recency effect describes the increase in recall at the end of the list. For example, serial position curves based on data, see Murdock Jr (1962) 30

2.1 Virtual Reality Examples: (Left) HTC Vive headset that is tracked in 3D space. (Right) A CAVE where the displayed image is projected onto a surrounding surface from the user's perspective. 44

2.2 Shows how immersive displays render to each eye on the same display creating stereoscopic vision. Graphics of the application are informative but optimised. 45

2.3 (Left) Shows how a mobile phone is placed as the display for Google Cardboard head-mounted display. (Right) Shows the placement of the Google Cardboard display in use. 45

2.4 Image sequence showing time on target selection method of camera icon (panel 1). The user reticule is positioned centrally to the viewport (Black circle, panel 1). The user positions the reticule over the camera icon (panel 2). Reticule expands to provide feedback that the icon is selectable. The reticule then redraws itself over 2 seconds (panel 3, white arrow shows the direction of redraw). When the circle is fully redrawn, the item is selected. 46

2.5 Users can teleport to the camera locations (Marked "C") at any point during the demonstration. The demonstration will not progress to the next movement until users select the "Next Step" icon symbolised by a white circle (Marked "S"). 48

2.6 Flow diagram to illustrate the sequence of events encountered by participants in this study 50

2.7 Chart to highlight participant response to "How would you describe your knowledge of the recovery position?" before using the training application. The majority (32) perceived their knowledge to be poor. 54

2.8 Box plot comparing recall of details using either visual and audio information or audio information alone. The median point of both audio and visual delivery is above the interquartile range of the audio-only boxplot. This suggests a significant difference between these groups as confirmed by a paired sample t-test 55

2.9 Graph to show recall for each step in the sequence of information delivered to users of the RPA. Error bars highlight significant differences between steps. Of interest, steps 1, 6, and 8 have the lowest recall. Generally, there appears to be a decrease in recall as the steps progress. 56

3.1 Flow diagram showing the series of steps for this experiment's methodology. 72

3.2 Room layout of the one-room (1R) (left) conditions snd two-room (2R) (right). S1-S4 = Speaker 01 - 04. Dotted lines indicated areas of visual interest. 73

3.3 (Left) Image of the two-room (2R) condition. Speaker is on the right of the image and the doorway in the centre with a view of the second room. (Right) Image of one-room (1R) condition. Both speakers are placed within one room 73

3.4 Boxplots of word recall across single-room (1R) and two-room (2R) conditions. 2R has a median value of 4 words recalled compared to a median of 5 for the 1R condition. However, the interquartile range and variance are similar, suggesting no significant difference between conditions. This is confirmed through significance testing 77

3.5 Graph to show the difference between recall at each speaker for each gender. The single-room condition contained S1 and S2, and the two-room condition contained S3 and S4. Error bars suggest that the difference in recall between speakers for each gender is not significantly different, while females significantly outperformed males at the free recall task. 77

4.1 Showcasing room layout of the single-room (Left) and two-room (right) conditions. S1-S4 = Speakers. The dotted line represents the navigatable area that a participant can explore in the virtual world 87

4.2 Above images show the speakers that deliver the word lists (left – Virtual World, right – Real World). Participants would tap the mobile phone screen in front of the speaker to activate the word lists. 87

4.3 Above floorplan shows the layout of the tutorial room. The dotted line represents the navigatable area a participant can explore in the virtual world. Here participants can test interaction with audio sources, using doors and throwing projectiles at targets to practice virtual controls and interactions. 89

4.4 Images to show visual aesthetics of the real world (left) and virtual world (right) for the single room conditions. 89

4.5 Boxplot to show the distribution of recall between conditions. RW1 had the highest average recall but also the highest variation. 91

4.6 Box plot visually shows the improvement in memory recall when using a memory strategy (Condition + MS). In real-world conditions (RW1, RW2) the effect of a memory strategy is greater than the virtual conditions (VW1, VW2), with least improvement to recall when using a memory strategy for VW2 compared to all other conditions. 93

4.7 The above line graphs show the recall per word for each speaker from those surveyed on whether or not they used a memory strategy (n = 16). The orange line represents those that used a memory strategy and the blue line those that did not. There is much more overlap per word recalled in the virtual conditions between those that used a memory strategy and those that did not. This suggests that the application of memory strategies is more consistent in the real world compared to virtual. 95

5.1 Factor diagram highlighting the four conditions for this study 103

5.2 Flow diagram to illustrate the sequence of events required to teach the player the required interaction and navigation paradigms, experimental condition protocol and recall capture. This sequence is split into a tutorial phase and an experimental phase to highlight the point at which the participants are expected to have sufficient knowledge to take part in the experiments. 104

5.3 Images of the tutorial phase levels. The top-left image is the level select room with the menu activation switch and iconography to walk towards it. The top-right is the interaction training room. The bottom-left is the experimental procedure training room, and the bottom-right is the recall room with maths questions as the distraction task. 105

5.4 These floorplans show the general layout of the experimental conditions. Even if the condition requires a single room or two rooms, the distance between the speaker is kept the same to help standardise the time between each speaker across conditions. 107

5.5 These four images show the four environments used for this study. On the left are the two single-room conditions, and on the right are the two-room conditions. Between conditions and rooms within each condition, the aesthetics and lighting are varied to foster the use of distinct environmental cues for each room and condition. 107

5.6 Image to showcase the six speakers required to add variation in visual attributes across rooms and conditions 108

5.7 Questions asked to explore perceptions of cognitive load during the experimental phase 111

5.8 Graphs to show regularity and context of VR use. Overall, participants had low experience with virtual reality, confined to other virtual reality experiments. 116

5.9 Boxplots for the number of words recalled per condition. 117

5.10 Boxplots that compares recall for those that used a memory strategy against those that did not for each condition. 118

5.11 Graph that compares mean recall for each voice used during word lists delivery for those that used a memory strategy. Voice M3 had significantly less recall than voices F3, M1, and M2 120

List of Tables

2.1 Table to show material delivered through the recovery position application. The steps indicate when the user can select to move on with the demonstration. The details per step describe the individual information elements within each step and represent the amount of material delivered at each segment of the sequence. The detail description synthesises the audio and visual information delivered for each detail. The modality delivery indicates if this information uses audio and visual representation, or audio only to deliver each detail. 53

2.2 SUS scores are used to categorise how useable and learnable the software is. P-values from the SUS scores are significantly above the mid-point of 3.00 from the Likert scale. This suggests a positive response to the usability questions. The overall SUS score is given with an appropriate adjective description relating to that score. 57

2.3 Table to highlight feature improvements and additions from interview discussions after using the Recovery Position Application on a mobile VR platform. 59

3.1 Participant Demographics 76

3.2 Post experiment survey 76

4.1 Shows that RW1 (Real World Single Room) facilitated significantly higher recall than all other conditions. 92

4.2 Shows that recall was significantly higher when a memory strategy was applied in all conditions except the virtual two-room (VW2). 94

5.1 Wilcoxon signed-rank test comparing each condition. No significant difference between the recall for each condition was observed. 117

5.2 Friedman tests across the counterbalanced voices used, speaker aesthetic, and word list. No significant impact on the recall was observed, suggesting that the word list used, voice heard, or form of the speaker did not significantly influence the recall of the word lists across the conditions. . . 119

5.3 SUS scores show sentiment towards the usability questions and how learnable the software is. P-values values suggest that the average response is significantly away from the mid-point of 3.00 in the Likert scale SUS survey. In this case. all responses are positive as they are significantly above the mid-point of 3.00. However, the high standard deviation of learnable questions suggests a neutral to strongly agree response in this area. 121

5.4 Mid-point value of the five-point Likert scale is 3.00. Significant mean values above 3.00 suggest agreement with the statement and below disagree. Results suggest that the tutorial rooms effectively taught the navigation and interaction paradigm. However, although significantly positive, there is wide variation in how automatic participants' interactions were with the virtual environments after the tutorial. 122

5.5 Mid-point value of the five-point Likert scale is 3.00 and therefore significant mean values above 3.00 suggest agreement with the statement and below disagree. Results suggest that verbally recalling the word lists felt initially "awkward", but that participants soon got used to this process. 123

5.6 Mid-point value of the five-point Likert scale is 3.00. Significant mean values above 3.00 suggest agreement with the statement and below disagree. Results suggest that cognitive load was focused on the memorising activity. However, there is some variation on whether or not the words presented were known to the participants and disagreement on if the VR worlds distracted participants from focusing on the words presented. 124

5.7 Mid-point value of the five-point Likert scale is 3.00. Significant mean values above 3.00 suggest agreement with the statement and below disagree. Results suggest that voices were clear and understood. Although a significant p-value was reached for the perception that voices were natural, there is some variation in this opinion. 125

Introduction

Motivation

If segmenting information between rooms aids the recall of this knowledge, can we use this as a design approach to improve the memory of information delivered in immersive virtual reality environments?

Virtual reality (VR) is a platform that can immerse users in synthetic worlds for various use cases, including entertainment, education, and research. The key to many VR experiences is how to pass information onto the user to aid knowledge transfer. This is most apparent when the platform is used for education and training. For example, before a learner can practice a sequence of actions, they will need to retain enough information about the movements for correct positioning and context of use. Content creators for VR experiences can choose any real or imagined environment to locate the user when exploring the VR material. The environments we inhabit when exposed to information to study can influence key processes of memory and learning. The environment can influence the cognitive load of the instruction (Choi, Van Merriënboer, & Paas, 2014), and shape how we interact with the learning content (Brooks, 2011). The environment can also act as a hook to the information contained within our experiences. When recalling past events, invoking the location of the experience can lead to a faster, more accurate recount (Hebscher, Levine, & Gilboa, 2018). How we separate information between locations (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016), and how we travel between these areas can significantly impact what is recalled (Smith & Vela, 2001).

Segmenting spoken content between two locations may aid the recall of this information compared to delivery within a single room. Using two 10-item word lists, Pettijohn et al. (2016) observed increased recall when participants listened to two word lists in different rooms compared to within one single room. It has been proposed that separating information between physical rooms can help to organise the word lists into two smaller lists that are easier to recall. From an event cognition perspective, this phenomenon has been explained by the generation of event models. An event model is a mental simulation of the events experienced within a given time and location (Radvansky, 2012). When recalling a day's events, we recall these event models and the associated experience. Since location is a key construct of an event model, we may likely develop an event model for each room we enter. By separating information between rooms and, therefore, between two event models, an individual is segmenting a larger body of information into smaller volumes that are easier to recall. In this fashion, the event models organise information on recall by chunking information and supplying individual cues for each volume of experiences.

VR is a platform that immerses the visual and audio senses of the user in a synthetic world. This world can be navigated with natural walking and can support interactions similar to the real world. Being virtual, it is also possible to move a user through various locations instantly whilst maintaining a psychological sense of being within the virtual reality and ignoring the real world. If information can be associated with an environment, then VR could use this principle to design learning spaces based on location change.

Research Questions

This thesis investigates the use of location change on recall within immersive VR and asks the following questions:

THEME 1: Can we utilise the segmentation of information between locations as an environment design approach to aid memory recall for VR training?

R1.1 What is the effect of segmenting information between immersive virtual locations on recall when individuals are not directed on how to encode or retrieve this information?

R1.2 What are the impacts of the visual and audio characteristics of the virtual location on recall when information is segmented between locations?

R1.3 How will individuals approach the encoding and retrieval of information when no specific strategy is required?

THEME 2: What are the challenges for replicating real-world experiments within VR?

R2.1 What is the impact of an immersive VR platform on cognitive processes which support the encoding and retrieval of information?

R2.2 How can we efficiently onboard users into a VR experience, so they have sufficient knowledge of navigation and interaction paradigms before experimental conditions?

Overview of experiments

A preliminary study explores the usability of the Recovery Position Application as a tool to teach the recovery position action sequence across mobile VR and desktop displays. This study establishes research methods for assessing learning outcomes from the application and participant perceptions of usability.

An important theme this work investigates is whether the segmentation of information between rooms can be used as a design approach for information retention within VR. To explore this theme, experiment one investigates if there is a relationship between the free recall of word lists and whether or not they are placed within a single or two-room VR environment. Experiment two investigates if there is a cognitive impact on recall when using VR rooms to segment information compared to real-world rooms. Experiment two uses a repeated measures design to have participants take part in the same memory activity from experiment one in both the real world and a VR world. A key challenge to preparing participants for a VR task with similar confidence to performing it in the real world is to provide the knowledge and practice for VR interactions and navigation. Experiment two also introduces a tutorial to teach and acclimatise participants to the VR world and its required interactions. Participants' perceptions of this approach are investigated. Experiment three investigates if recall from segmented information between VR rooms is mediated by variation in word lists' visual and audio characteristics. In this experiment, participants are tasked to recall from two word lists placed in one or between two rooms. The visual look and the voice used are either kept the same or different between these word lists. Within VR, contact with any real-world entity may remind the individual of the real world and break immersion. Experiment three iterates on previous VR tutorials as it is designed to be completed with limited contact with the experimenter. The tutorial aims to teach interaction and navigation techniques and the experimental procedure. This tutorial is assessed through the perceptions of usability, participants' confidence in achieving the required tasks, and perceived cognitive load. Through experiments one to three, participant approach to memory tasks is analysed to see if an individual's approach overshadows any use of environmental cues to organise the recall of word lists.

Overview of results

The preliminary study highlights that virtual worlds across different displays and user interaction can establish a mental representation of an action sequence. This affirms that virtual worlds can effectively deliver spatial information like body movements. Experiment 1-3 evidence that segmenting word lists between immersive VR rooms does not mediate recall. However, there was a confounding variable of memory strategy use. Experiments 1-3 suggest that the VR tool effectively teaches interaction, navigation paradigms, and experimental procedures. Additionally, the implementation of the experimental conditions is fit for purpose.

Contributions

1. Across experiments 1, 2, and 3 of this thesis, we evidence that segmenting word lists between VR rooms does not mediate free recall. This work suggests that when delivering verbal information within a VR space, segmenting this between locations will not aid retention. The methodology used to explore this interaction was based on previous real-world studies that have observed an increase in recall (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016; Smith & Vela, 2001). These results, therefore, contradict this work when explored within an immersive VR context and develop an interesting line of enquiry to explain this difference. To the author's knowledge, this is the first attempt to replicate this type of methodology within VR by simulating virtual locations that use the navigation paradigm of physical walking.
2. When replicating Pettijohn's methodology Pettijohn et al. (2016) for assessing recall of segmented word lists between rooms, we observed a significant effect of memory strategy in both the real-world (experiment 2) and VR variants (experiments 2 and

3). This suggests that this methodology facilitates the use of memory strategies that may supersede and obscure any room segmentation effect on word list recall. In experiment 1, the use of memory strategies did not significantly mediate recall. However, due to the strength of the effect in experiments 2 and 3, the lack of effect in experiment 1 may be due to participant variability in applying the memory strategies between study cohorts. The use of memory strategies highlighted in experiments 2 and 3 is a significant confounding variable not accounted for in previous work by Pettijohn et al. (2016). Recommendations for mitigating this confound and gathering appropriate metrics are suggested for future work.

3. In experiment 3, visual and audio variations in information delivery did not significantly affect word list recall. This suggests that whether or not participants choose to use a memory strategy to aid recall, local audio/visual variation of the information delivered does not significantly mediate cognitive processes used in the subsequent free recall of word lists. These results suggest that when designing VR spoken content, the visual and audio characteristics of the information delivery can be kept the same across locations without significantly impacting users' encoding and retrieval processes.

4. The approach to onboarding participants into the VR world and experimental conditions was analysed and subsequently iterated between experiments 1, 2, and 3. Therefore, the tutorial method in experiment 3 can inform good practices for VR onboarding. This approach would slowly build up the skill set of the participants within the immersive virtual world by establishing basic interactions and allowing practice before adding more complex variants. This is repeated until the entire navigation and interaction paradigm is sufficiently practised. Participants would then take part in a mock experiment to meaningfully apply what they have learnt and understand the rules of engagement for the study. In experiment 3, the participant perception of their cognitive load was low for the navigation and interaction paradigm, with high confidence in their ability to interact with and navigate the VR world. Cognitive load was perceived to be induced by the memory activity, not the VR platform. Based on this work, we can highlight future directions to improve onboarding when using VR technologies for research.

Chapter 1

Background

1.1 Chapter Summary

A preliminary study (See chapter three) was conducted to evaluate the viability of the mobile virtual reality application, "The Recovery Position Application". From this study, the question was asked, how can we improve information retention within virtual reality training? The motivation for this question is to guide the design of immersive virtual reality environments for education and training. The literature explored in this background section explores the use of information segmentation between immersive virtual reality locations as a potential design approach to facilitate better information retention.

The background chapter has three main sections that explore:

1. **Can Location Change Mediate Recall?** - Highlights research that connects theories of event cognition, environmental context, and location to recall.
2. **Serial Position Curves of List Recall** - Describes common attributes of word list recall in humans, as this is the dependent variable in many of the studies in this thesis
3. **Virtual Reality Affords Spatial Cognition** - Outlines how immersive virtual reality could afford similar spatial cognition to the real world and therefore be a suitable platform to apply theories of information segmentation between locations on recall

1.2 Can Location Change Mediate Recall?

Recalling the experiences of each day is part of how we learn, reflect upon and communicate our behaviours in a variety of scenarios. This daily stream of detailed information can be described as a series of events, a set of experiences given context by the chronology of order, actors and location. Understanding the mechanisms that enable a retrospective of the day's experiences can highlight automatic processes of cognition for the recall of events as well as information within these events.

Mental models describe a broad set of cognition that creates "structural analogues of the world" (Johnson-Laird, 1983). An internal representation of reality that enables us to mentally simulate, learn and better adapt to live within our external environment. They provide an understanding of external systems so that we may predict outcomes, influence, and empathise with these dynamic structures. These cognitive representations of systems and situations are accessible and long-lasting (Furlough & Gillan, 2018). **Event models** refer to a sub-type of mental model. Event models are the mental representation and simulation of events (Radvansky, 2012). An event can be described as an amount of time within a location that is perceived to have a beginning and an end (Zacks & Tversky, 2001). Therefore, the crucial constructs of an event model are the spatial location and time of day that an event takes place. For example, if an individual were to recall what they ate for breakfast, they may first bring to mind broader location details like the room or try to picture more local environment details like the table or chair that was used during this event. Event models act as a structure to recall both the chronology and content of information that an individual is exposed to.

Event models give access to the context of experiences, as well as provide a hook to the information exposed within each event. Since event models are thought to be accessible and long-lasting, some work has explored how the transition from one event to another can influence the recall of information observed within an event. The point at which one event model ends and another starts is called an **event boundary**. (Flores, Bailey, Eisenberg, & Zacks, 2017) whilst watching a movie of individuals performing everyday tasks, participants were asked to try and remember as much as they can from these films whilst pressing the space bar on the PC every time they felt they came to the end of an event. participants were asked to recall these videos at different time delays of 10 minutes, one day, one week, and one month. When pressing the space bar and therefore emphasising their own event structure towards these watched videos, this group recalled significantly more across all time delays compared to groups that were asked to try and remember as much as they could. This suggests that emphasising own event structure and boundaries around an event may help recall across short and long time delays. However, when asked to recall straight after the videos had been presented, there was not a significant difference between those that emphasised the events they saw and those that did not.

Creating own interpretation of events may enhance recall compared to imposing an event structure. Again when showing movies of everyday activities (Gold, Zacks, & Flores, 2017), event boundaries were emphasised by slowing down the video and playing a bell sound. This significantly increased recall of event boundaries but not the information within the event. This may suggest that event structures can be shared but associating information with these structures requires some engagement from an individual.

A key construct of an event model is the location with which the event is associated. Although the boundary of where one event model ends and another starts can be established by any significant focus of attention within a location, noticeable changes in the environment are thought to define the placement of event boundaries. By walking between two locations, it is theorised that an individual will generate an event model for each local, for example, location changes by walking through doorways. Since event models are theorised to structure the recollection of experiences, research has examined the effect of location change on an individual's ability to recall information from past events. Research suggests that walking through doorways may hinder and enhance recall, depending on how information is organised between event models.

Associating the same information with multiple event models may reduce the recall performance of that information. In a study using virtual environments, individuals were tasked to carry an item between two virtual tables and pick up a new one (Pettijohn & Radvansky, 2018). The tables were either in the same room or between two rooms. Recall of what item was in hand or just placed down was probed. Notably in a two-room condition, the object in hand would represent the current event model, and the object just placed down would represent recalling information from the previous event model. Recall performance reduces when objects are associated with two event models. The rationale is that when recalling from two event models, there is competition between the two which reduces performance.

Segmenting information between event models may have the opposite effect and enhance recall. Pettijohn et al. (2016) explored this concept in three different contexts. Firstly, participants were exposed to two 10 – item word lists read out by the experimenter. After both word lists were delivered, the participant would take part in some maths questions as a distraction task and then immediately after being asked to recall on paper as many words as they could remember. On one condition, these word lists appeared in a single room. In the comparison condition, these word lists were placed in two rooms with a doorway in between. For both conditions, the distance of the word lists was the same. participants took part in both conditions and recalled significantly more when the word lists were segmented between rooms. In the second experiment, participants were shown word lists on a desktop monitor in either one window or between two windows. After a distraction task, participants typed down as many words as they could recall from the word lists. Using two windows observed significantly higher recall rates. The third and fourth experiments had participants read narratives. Within these narratives, the number of spatial changes was changed for each condition. participants had significantly better recall when there were more spatial changes within the story texts. This set of experiments suggests that a location change in the real world, in the local space of a PC screen, and the mental simulation of a narrative can aid recall of word information presented.

The apparent contradiction that walking through doorways can both aid and hinder recall is the subject of the Event Horizon Model (Radvansky, 2012). This model was developed as a means to describe the behaviours observed when individuals pass through event boundaries. This model states five principles:

1. Individuals segment the stream of daily information into event units – These units are stored in memory. When an individual reaches an event boundary (new location, significant focus of attention), a new event unit is created. This process is automatic and based on our focus of attention
2. The superior availability of information in the working event model – The working event model describes the event that is in working memory, i.e. being experienced. Information relating to even prior event models is less accessible than the current one.
3. The construction of a causal network that can then influence retrieval – the causal

structure of events helps aid the temporal understanding of the sequence of events. It is easier to process material that is causally connected.

4. The superiority of memory for information stored across multiple events in non-competitive attribute retrieval. – Segmenting information between event models can help organise and chunk this material. The event models have their own mental structure of connections that act as containers for the information we experience within. Therefore a larger body of information can be segmented between these event models and be more easily retrieved.
5. The occurrence of retrieval interference for information stored across multiple events in competitive event retrieval – if multiple event models share that same information, then they will all be retrieved when recalling this material. Therefore there is competition for cognitive resources as multiple event models are being scanned for the wanted information.

The event horizon model explains how walking through doorways can both aid and hinder recall through principles four and five. If an event model is created for each location entered, then separating information between rooms associates the segmented parts of the information with individual event models. These event models are essentially chunking a larger body of information into smaller volumes. Memory research has shown that organising information into "chunks" improves retrieval (Mandler, 1967). A "chunk is a collection of elements having strong associations with one another, but weak associations with elements within other chunks" (Gobet, Lane, Croker, Cheng, Jones, Oliver, & Pine, 2001). For example, to help remember the list: rose, duck, petal, bloom, robin and seagull, an individual can associate rose, petal and bloom to the category "flower", and duck, robin and seagull with the category "birds". At recall, the category or rule is remembered and aids retrieval speed and accuracy of the associated information, making working memory more efficient given capacity limitations (Bower, Clark, Lesgold, & Winzenz, 1969). Even when no structure is provided, an individual may organise information as best they can to aid retrieval (Bousfield, 1953). When information is associated with many locations, then multiple event models are recalled for a single piece of information. This creates a **fan effect** which describes a reduction in memory performance or increase in error rate for a concept as associations increase (Radvansky & Zacks, 1991). An example would be recalling the last place you may have left your keys. This may recall a variety of locations that you then have to think through to recall the correct location, taking longer to parse through and perhaps needing to check each location until the correct placement is identified.

Similarly to event cognition work, environmental context-dependent memory research has also found evidence of improvements to recall of word lists when encoding occurs between multiple room locations (multiple contexts). Smith (1982), had participants learn four word lists in either a single room, between 2 rooms, or between four rooms. Participants would wait in a hallway between location changes. Recall significantly increased as room numbers increased. However, this result was not replicated in future work with a similar methodology (Smith, 1984, 1985), where participants learnt Lists across three rooms. Smith interprets the data that a trend in increased recall was observed; however not significant. In a meta-study exploring environmental context on memory across several paradigms, Smith & Vela (2001) found that using a multiple environment paradigm to mediate recall had a modest and reliable effect size (d = .45). Smith did find that when lists were learnt in a single room, and participants were asked to think about the room environment on recall, a significant increase in recall occurred (Smith, 1984). This suggests that incidental cues from the environment can aid recall when used but may not always be used by individuals. When using MRI scans to analyse brain volume activation during episodic autobiographical recall (memory of events that happened to self), Hebscher et al. (2018) found similar results. Using location early in the recall of past events did correlate to faster and more accurate recall, but there was a large variation in the approach taken by participants. In line with these findings, Robin et al. (2018) used fMRI scans to observe the neural firing when recalling locations, objects and people. The neural representation of location was seen across a large set of brain regions, compared to objects and people. The event representation accuracy significantly increased when based on location compared to objects and people.

How information is either segmented or shared between event models may mediate recall. Event boundaries can be enhanced by some meaningful interaction but this does not correlate to more incidental knowledge of experience within that event. Location does appear to be a fundamental construct of events and also a hook to information within. If location is a crucial construct of an event model, then how we navigate a series of environments may act as a bridge between event structure and the information experienced within each event. When individuals explore an environment, walking through doorways can both hinder and enhance the recall of objects found within. By segmenting information between locations, there is the potential for increased recall without the aid of a specific strategy on the part of the participant. This is supported by work within the environmental context domain. However, inconsistencies in replication suggest that individual approach to recall may be a strong influence on whether environmental cues are used or if an individual is practised at using environmental cues to aid recall when recalling events.

1.3 Serial Position Curves of List Recall

When exploring the effects of segmenting information between locations, one approach is to have participants attempt to learn a word list at each location (Smith, 1982; Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016) and then administer a free recall test. Ideally, the recall location would be in a neutral location to avoid the confounding variable of **context reinstatement**. Context reinstatement refers to an observed increase in memory when recalling information in the same environment that which it is encoded. This effect is thought to derive from individuals being able to use memory hooks from the environment that are incidentally associated with the learnt content of that local. A classic example of this is where Godden & Baddeley (1975) had participants learn word lists underwater in SCUBA dive gear and on land. Recall of these word lists was significantly higher when performed in the environment where encoding occurred. When using a neutral environment for list recall, an individual cannot leverage the context reinstatement effect, which could overwrite the use of any cues that are the subject of the study.

Word lists have been used throughout memory research. They are a practical body of information that can be analysed for frequency of recall, order of recalled words, and associations between each word. From extensive use across studies, some robust patterns of word list recall in humans have been identified. A commonly observed pattern is the serial position curve. This describes a "U" shaped curve when plotting the recall accuracy of a word list against the item order of the words delivered (See Fig. 1.1). This serial position curve is expected when the word lists delivered are unstructured and followed by an immediate recall test. The unstructured description means the words are random, equally common in everyday use, and don’t have any explicit association between them. As the length of word lists increase, more words will be recalled, but this will account for a diminishing proportion of the total words (Murdock Jr, 1962). The "U" shape is explained

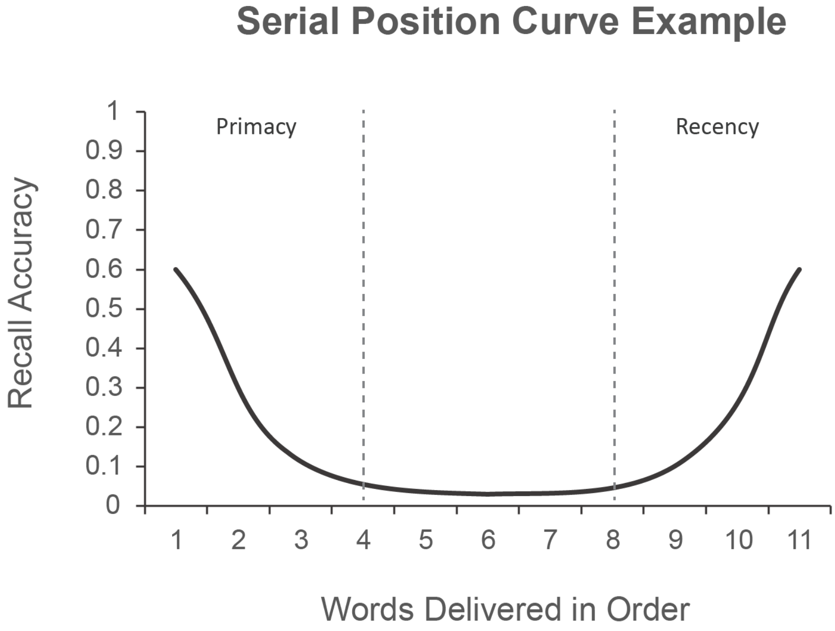

Figure 1.1: Example serial position curve for descriptive purposes only, and not based on a specific data set. This illustrates the primacy and recency effects of word list recall. The beginning and end of a delivered word list will be recalled more often compared to the middle items when the list is unstructured, and the recall test occurs straight after delivery. The primacy effect describes the increase in recall at the beginning of the list and the recency effect describes the increase in recall at the end of the list. For example, serial position curves based on data, see Murdock Jr (1962)

further as primacy and recency effects. The primacy effect describes how the initial portion of word lists are commonly recalled and is considered a strong effect. Potentially, this is due to having more time within cognitive resources to form long-term memories. Additionally, when a list starts, these words will be more easily the focus of attention with the most amount of cognitive resources spare. Essentially creating the best environment for long-term encoding. As the word lists progress, words will be presented faster than can be encoded in long-term storage. Therefore, the words will need to be held in working memory, reducing available cognitive resources and splitting attention between listening to the new words and encoding the previous words. The recency effect describes the increase in recall for the end portion of words delivered. When a recall test is administered straight after the word list is delivered, the recency effect is caused by an advantage for the words held in working memory and therefore, may not be a sign for long-term storage of these words. This is highlighted by simply adding a time lag between word list delivery and the recall test. A 30-second time lag is enough to significantly reduce a recency effect (Glanzer & Cunitz, 1966)

When more structure is added to a word list, the primacy effect rises above the recency effect (Deese & Kaufman, 1957). As the unstructured words turn into an understood sentence, then the primacy effect is shown not just on the individual words but also on the narrative between them. Adding structure to a set of words will make them far easier to recall. In the study of mnemonics, finding ways to categorise and find relationships between materials is the aim of many memory strategies. If one then recalls the category or mental "hook" to a category, one can then recall the associated information. Persensky & Senter (1969) observed that when participants associated listed words with their own mental "hooks", significantly more words were recalled. When looking at the serial position data, this recall still showed signs of primacy and recency effects but with a much higher proportion of middle words in the lists being recalled. The memory strategy therefore, significantly flattens the typical "U" shape of the serial position curve. In this same study, when a delay was placed on recall, then the recency effect again diminished. A common mnemonic for word lists is to create a story from the words to aid recall. When participants were asked to make a narrative from presented word lists, a far greater recall was observed when a participant created their own story compared to listening to another. The serial position curve was still present, including the recency effect. This could be due to the end of a narrative having more cognitive impact than the end of a word lists (Herrmann, Geisler, & Atkinson, 1973). The serial position curve is a robust effect on the recall of unstructured word lists. It is also a good baseline to discuss the effects of individual approaches to recall from word lists. Studies have shown that participants who use memory strategies will flatten the serial position curve but still show signs of primacy and potentially recency effects.

1.4 Virtual Reality Affords Spatial Cognition

Extended reality is an umbrella term for immersive technologies that merge digital realities with the real world for human participation. These technologies sit on a spectrum from augmented reality, where virtual content is overlaid in the real world, to virtual reality (VR), where the participant's reality is replaced by a virtual one (Milgram & Kishino, 1994). Virtual reality is a computer-generated synthetic world which physically immerses the participant's senses to create the illusion that the user is in a new reality (Slater, 2014). The immersion describes how a participant's visual and audio feeds are replaced by virtual ones. A standard, commercial approach is through a head-mounted display (HMD). When a participant wears the HMD, they will have a screen positioned close to each eye, and speakers close to their ears. As the participant moves their head, the HMD is tracked in 3D space and updates the sights and sounds accordingly. This tracking uses six degrees of freedom (6-DOF) which means that both the rotational values and positional values of the headset are tracked. This gives the impression that as the user moves their head, they are moving it around a virtual world. The lens for each eye enables the replication of a stereoscopic view with human perspectives Slater & Sanchez-Vives (2016). The VR hardware facilitates a realistic perspective, audio and visual feed, and movement within 3D space.

Through game engines, the content of these virtual realities can be developed. A game engine is a development environment with core libraries and functionalities traditionally aimed at creating games. Although the features vary between game engines, they commonly include a solution to render virtual content, animate characters, particle simulations, physics simulations, and logic to trigger events based on user interactions. An example game engine would be Unreal Engine 4 ue4. Although traditionally aimed at creating games, game engines are now used across many more industries. For example (but not limited to) education, industry training, architectural/world visualisation, research, and film production. These engines can create virtual worlds with realistic lighting, materials, physics, and real or imagined characters. As a user, you can interact with content and events, explore, and take part in the narrative of the world. Using a VR platform to interface with game engine worlds means that a participant can be exposed to synthetic experiences from a human perspective.

VR platforms also afford interaction in these virtual worlds. This is generally through 3D-tracked controllers in the participant’s hands. Through these controllers, objects can be picked up, doors can be opened, and items used similar to the real world. The combination of interaction, the reaction of realistic physics, and sensory attributes create the illusion or plausibility that the virtual world is now the active world for the participant (Slater & Sanchez-Vives, 2022) and engenders a sense of "presence" (Witmer & Singer, 1998). Presence describes the psychological sense of "being there". The acceptance of the illusion of VR that the attention of the participant is in the new virtual reality, and ignoring the real world (Slater (2014)). A sense of presence has been a defining attribute of VR. The psychological acceptance of the virtual reality means that a participant is not just engaged attentionally with the world but believes they are part of the virtual reality, not just an avatar they control. The acceptance of the VR illusion and physical immersion of the senses means that VR can be used to simulate experiences and cognition of the real world for at least a short amount of time.

Spatial awareness and sense of scale require knowledge of objects in relation to self and each other. An egocentric spatial frame describes the cognitive understanding of the environment in relation to self. For example, "Environment object is X distance away from me" or "Environment geometry is X times bigger than me". An allocentric spatial frame describes a cognitive understanding of the spatial relationships of objects to each other. Whilst observing an environment, both egocentric and allocentric information is implicitly analysed. Through these frames of reference, we can build up a cognitive map of an environment to make choices for actions like navigation. Egocentric and allocentric cognitive representations of space may work in tandem to understand our immediate spatial environment and establish memories of the local (Burgess, 2006). The interplay of allocentric and egocentric spatial frames can be viewed as dynamic with the creation of allocentric spatial frames from egocentric ones and vice versa (Ekstrom & Isham, 2017). As the knowledge of one reference frame can inform and augment another, the accumulation of many spatial frames can therefore help develop the memory of an environment and objects that are placed within it. For example, viewing an area from multiple vantage points. (Kelly & McNamara, 2010). The perspective of VR is close to the natural, egocentric perspective used by humans in everyday viewing of the real world. This egocentric perspective, immersion and sense of presence can therefore afford spatial understanding similar to the real world. The combination of spatial understanding, and virtual content that can simulate real or imagined environments and experiences, makes VR an exciting platform for spatial cognition research (Diersch & Wolbers, 2019; Stangl, Achtzehn, Huber, Dietrich, Tempelmann, & Wolbers, 2018).

The physical immersion of the visual sense may also aid spatial memories of objects and information explored within a room compared to non-immersive displays like a desktop PC. Ragan et al. (2010) had participants learn the sequential positioning of shapes in an immersive virtual environment (CAVE). Between groups, their level of immersion was varied by adjusting the angles for the field of view (the angular area that can be viewed at any given time)and the field of regard (the number of CAVE screens that were projected upon). Higher levels of immersion resulted in better memorisation of the task and this transferred to the real-world version of the task. This has extended to mnemonics that use location as the framework for later recall. When cave experts were asked a series of questions based on a 3D virtual map of a cave (allocentric representation, tasks that involved finding map details and comparative measurements were answered more quickly, and with greater accuracy using higher immersive displays (larger field of view, stereoscopic rendering turned on) Schuchardt & Bowman (2007). When exploring memory places, Krokos et al. (2019) found that when participants were tasked to associate face-name pairs to loci within a virtual world whilst being seated, using an HMD compared to a desktop display with a mouse for navigation resulted in significantly higher recall for this task.

However, there may be some differences between the spatial processing of environments when using VR compared to the real world. In a study by Kimura et al. (2017), participants were placed in a wheelchair in both an immersive VR condition and a real-world condition. When in the wheelchair, participants were disorientated by being blindfolded and moved around the room. When the blindfold was taken off, they were asked to find the right corner (based on a training phase). Conditions were varied, so participants had to rely on a variety of visual cues to re-orientate themselves. The study found that participants used the same cues in both the real world and virtual, but the weighting was different. Geometric cues were not encoded in the virtual world as much in the real world. This indicates that although general spatial processing is similar between realities, there is nuance in the degree to which these cues are utilised. Therefore, studies that use more nuanced elements of spatial processing may find it difficult to observe transferable findings between real-world and VR scenarios.

VR affords many types of navigation through a virtual world, from sitting down and using a joystick to translate position to natural walking. How we navigate through an environment may be a mediating factor in our spatial understanding. Zanbaka et al. (2005) had participants explore a virtual room for five minutes in one of four conditions: VR with natural walking, VR HMD with the ability to stand whilst translating their movement with a joystick, VR HMD with just the rotation tracked and a joystick to navigate, a desktop monitor with a joystick. After the exploration, the participants were asked to take part in a cognition questionnaire and sketch the virtual room. The results from the questionnaire imply that the ability to explore a virtual environment naturally might be beneficial for situations which require problem solving, interpretation, synthesis, or evaluation of information. There was also a positive correlation between the questionnaire and sketch map scores. This may indicate that natural walking through an environment can better aid the cognitive map of a virtual room. Similar results were found by Ruddle et al. (2011). This study explored route finding in immersive VR environments. Those that physically navigated through the routes made fewer errors than those who had natural rotation but required a joystick to translate their position through the environment. These studies may highlight the vestibular system as a key source of information to develop spatial understanding or that the alternative forms of navigation in these studies are less practised and therefore, more difficult to develop incidental spatial knowledge. It does suggest that using natural navigation (walking) through an immersive VR world will afford better spatial understanding.

1.5 Challenges When Using VR For Experimental Trials

Presence is a key design consideration when using VR for experimental studies because if users lose their sense of presence, then they will disengage from the virtual world presented. As long as presence is established, it is more important to make sure presence is not broken. If it is, then an individual will be reminded of the real world around them, start to process both the real and the virtual world, and begin to disengage their attention from any task within the virtual realm.

There are a few factors to consider when designing a virtual world that does not break presence. Simulation sickness (SS) (Brooks, Goodenough, Crisler, Klein, Alley, Koon, Logan Jr, Ogle, Tyrrell, & Wills, 2010) can brake presence by inducing the feeling of nausea. Additionally, any sense of nausea could interfere with the memory tasks asked of the participants. On rare occasions, virtual environments have been known to induce SS, which can cause symptoms very close to motion sickness but less severe. SS is individual to everyone, and most do not experience it at all. If you are currently ill (feeling nauseous), then there is a higher chance that you will suffer from the effects of SS. To reduce the risk of SS, a minimum recommended 72 frames per second for interactive applications (Developer, 2022), and kept stable. Although other factors like vertical movement in virtual space can cause SS, they will not apply to this experience as the user will navigate through walking and is not having their sense of movement altered by the experience.

Another key element to maintaining presence is to have features and interactions act as expected and remain bug-free. Any interaction that appears broken or abnormal for the world may remind the user that they are interacting with a fallible system and not a reality. A participant must be sufficiently trained on the interactions and experimental procedure before taking part in a VR trial. If a participant does not know what they need to do when working through the experimental conditions, they will want to ask the experimenter about what they are required to do. Since the experimenter is not within the immersive virtual world, then this conversation may remind the participant of the real world and break the immersion of the virtual world. To maintain presence, it is therefore important to limit any contact with the real world during experimental conditions.

Using VR for an experimental study will require some form of training for the participant before the experimental conditions. Participant experience with VR cannot be relied upon for a consistent knowledge base of navigation and interaction paradigms. Although virtual reality has increased its commercial accessibility, it is not ubiquitous across many population samples. There is variation in hardware and software features between headsets and much ongoing development. Therefore, even if many participants had experienced virtual reality previously, it is most likely that this experience differs greatly between each participant. The speed of adaption to the controls of the virtual world will largely depend on the previous experience of the user. For example, if a participant has had previous experience with virtual reality, they will most likely understand the controls more quickly than someone who has not. Therefore, any extra cognitive load devoted to actively thinking about how to interact is increased in those who have less experience. Additionally, if a participant is thinking about how to interact, then their attention is divided between the task and how to interact with the task. Dividing attention when encoding information can significantly reduce the ability to recall (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996).

By the time a participant takes part in the experimental conditions, they need to be comfortable with the navigation and interaction paradigms, understand what they need to do within the experimental conditions, and be acclimatised to the VR hardware and virtual representation of self. Therefore a tutorial phase is required before the experimental conditions to train and acclimatise participants consistently across the population sample.

Tutorials are important before engaging with any virtual world. A tutorial is a form of instruction that enables a user to play through a virtual experience by teaching information on interactions and the context of the virtual world. In the entertainment space, computer games provide good examples of effective tutorials. Tutorials for games need to be engaging and swiftly allow the player to effectively play the game at a beginner level. A common approach is to create a level that guides the player through a set of interactions whilst highlighting how these relate to the win-and-lose conditions. Importantly, one interaction is explored at a time, practised, and then built upon through the tutorial level. For example, if a game requires a player to jump over obstacles, then an effective approach to teaching how to jump would be:

1. Instruct the player to press the "jump" button to provide knowledge of the interaction
2. Give the player something to jump over or onto to progress. This reinforces where the jump button is and adds context to its use without yet penalising the player
3. Now, have the player jump over a gap. This challenge has more risk and requires knowledge of how far the player can jump. Completing this challenge would signify some good progress has been made with the player's understanding and efficacy through the practice of the jump mechanic. If they fail, they can restart close to the challenge and practice until they are proficient enough to progress.
4. We can now move on to the next mechanic, which can be new or built upon the jump mechanic.

Importantly, in this example, the player is engaged with the game world but also learning incrementally the required interaction and navigation paradigms that will be used as a basis for the gameplay to follow. Additionally, the challenges presented facilitate the practice of a mechanic, so automaticity of use is fostered. This means the player does not think about how to jump, this interaction is now automatic. Instead, they are thinking about when to jump concerning the challenges presented.

If a participant does not understand how to automatically interact with the virtual world, then cognitive resources will be used to actively think through each interaction. This would take attentional and working memory resources from the task at hand by increasing extraneous cognitive load.

1.6 Summary

VR can support spatial cognition similar to the real world by providing an egocentric perspective, a world to explore, interactions that respond as expected, and natural navigation through walking. VR also affords a pragmatic environment with great control over the spatial variables within a single laboratory. These attributes make VR a compelling platform to apply theories that connect the spaces we inhabit to our memories. If walking through doorways can indeed improve memory, then walking through immersive virtual doors may observe the same effect. The advantage of VR is that it is practical to transition between multiple environments whilst being in the same real-world location. Therefore location transitions have the potential to be used as a design philosophy to aid the memory of information delivered.

Chapter 2

Preliminary Study

Exploring the use of mixed reality displays for observational learning of motor skills: A design research approach enhancing memory recall and usability

2.1 Chapter Summary

The Recovery Position Application (RPA) was developed (ISS, 2016) within the Interactive Systems Studio at the University of Plymouth. The application aims to test the viability of using mobile virtual reality as a platform for learning the recovery position procedure.

This study was conducted using RPA to gain experience with experimental methods that evaluate virtual reality training systems, their design, and learning outcomes. In particular, methods that analyse usability, presence, and knowledge transfer: This chapter can therefore be viewed as an extension to the background of this thesis. Additionally, this study was used to frame thoughts on thesis motivation and direction by reflecting on what parts of the information delivery could be explored to aid knowledge transfer.

The introduction and background of this study outline why virtual reality is a suitable environment for spatial information like movement analysis. It then justifies the system's design concerning motor skill acquisition, demonstration-based training, and application interaction.

2.2 Motivation

When learning an action sequence, observing a demonstration informs the knowledge of movement execution and enhances the efficiency of motor skill acquisition. 3D virtual learning environments offer more opportunities for motor skill training as they afford observational learning. Mixed reality platforms (virtual reality, desktop PC etc.) that render 3D virtual environments can therefore increase the accessibility of observational content. To explore these platforms' effectiveness in facilitating observational learning of action sequences, we developed the Recovery Position Application (RPA) (ISS, 2016) at the Interactive System Studio, University of Plymouth. The RPA was initially designed for mobile virtual reality. The RPA displays two virtual avatars performing the steps of the recovery position. We present the design of content and interaction informed by research into observational learning of motor skills. To evaluate the current functional prototype and its potential use within an educational context, RPA was tested on three different platforms. Mobile VR (N=20), desktop display (N=20) and video recording (N=21). Memory recall of movements was recorded, and the usability of the RPA was investigated. Across all three platforms, the average recall of demonstrated information was 61.88% after using the application for 10 minutes. No significant differences in recall rate were identified between platforms. Participants' responses were positive for both application effectiveness as a learning resource and ease of use. These results are discussed regarding the future development of the RPA and guidelines for virtual demonstration content.

2.3 Introduction

In training and education, there are many instances where students will need to imitate a performance from a demonstrator. For example, to understand how to use laboratory equipment and computer software, or acquire motor skills for sports, etc. Demonstration-based training (DBT) (Rosen, Salas, Pavlas, Jensen, Fu, & Lampton, 2010) requires effective delivery of observational content for students to learn. A demonstration is a "dynamic example of partial – or whole-task performance" that conveys the learner's required knowledge, skills, and attitudes. The two learning opportunities are when the student observes the demonstration and when any activity supplements the understanding of this performance either pre, during, or post-demonstration.

DBT is a common approach used to teach motor skills. For example, in acquiring a set of dance movements, the teacher will demonstrate an action and then ask the student to imitate said action for practice. Central to this process is the use of observational learning by the student. Although physically practising a motor sequence grants an implicit long-term memory of movements (Boutin, Fries, Panzer, Shea, & Blandin, 2010; Wulf & Schmidt, 1997), the addition of observation enhances the efficiency of motor skill learning (Ashford, Bennett, & Davids, 2006). Through observation, an individual can acquire a mental representation of a motor skill to cue imitation (Sheffield, 1961) and correct errors. Fitts' and Posner's (1967) model for learning motor skills describes three typical stages: cognitive, associative, and automatic. This model establishes that cognitive representation is important at the beginning of motor skill development, where knowledge of movement positions and goals are limited. Later into motor skill development, the learner may still benefit from more demonstrations as they refine their movements, but applying the technique and feedback becomes more critical to learning. Knowing how to execute an action does not mean an individual is proficient at performing said action. A student will inform their progress throughout training with self-analysis and instructor feedback. Through sleep, cognitive representation and motor neuron information from physical practice will consolidate (Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002). Therefore, students will typically need to practice to develop a motor skill over many days and weeks. Applying a motor skill to various scenarios will develop generalisability of use. Identifying when a student needs to vary their training or focus on a specific detail is informed by the goals of the student and the judgement of the instructor (Williams & Hodges, 2005). Demonstrations aid the process of motor skill development by providing a mental representation of actions to inform movement goals during practice sessions and information to help define criteria for feedback. Therefore, the timing and content of a demonstration will depend on the structure of a training program and the current level of experience of the student with a motor skill.

If cognitive representation is an outcome of observing a demonstration, the representation will be encoded into memory (Fitts & Posner, 1967; Bandura & Walters, 1977). This memory may be symbolic and subject to decay (if not practised or rehearsed). However, it will provide information for the user to decode, interpret and subsequently imitate by providing familiarity and valuable analysis not available while performing the actions

(Bandura & Walters, 1977; Elliott, Grierson, Hayes, & Lyons, 2011).

3D virtual environments with animated avatars extend the opportunities for observing demonstrated content inside and outside the classroom, affording realistic spatial knowledge representations by replicating real-world perspective and lighting (Dalgarno & Lee, 2010). Desktop PCs and mobile devices, like tablets and smartphones, provide a variety of platforms to present virtual demonstrations. Immersive technologies like head-mounted displays (HMD) can visually and audibly envelope the user as if they were in an actual environment, mirroring a real-world viewpoint. This viewpoint is egocentric (displays objects in relation to the user), which will aid the general mapping of environments, including allocentric representations (objects in relation to each other but not to self) (Epstein, Patai, Julian, & Spiers, 2017). Yearly advancements in smartphone performance have increased accessibility to devices that can render 3D virtual environments. The improved functionality has enabled smartphone-focused virtual reality (VR) platforms like Google Cardboard (Google, 2022) that have been widely adopted for entertainment. Mobile VR has significant potential to aid observational learning by providing virtual demonstrations inside and outside the classroom. This form of demonstration supports in-class training, facilitates DBT at long distances and enables observational learning for independent study.

To investigate the use of virtual content as a tool for observational learning of motor skills, the Recovery Position Application (RPA) was developed for the Google Cardboard platform. This application aims to show the recovery position action sequence for the user to observe and memorise. Within the framework of DBT, this application has one function: display the demonstration. To apply relevant theory to the design of the RPA, a software development approach was used. The software development design process establishes the software's requirements (needs) based on the business case (problem domain). These requirements are then broken down into features and functions that consider the technology's target audience, software and hardware. Once developed, these features are then tested to evaluate their implementation. The critical requirement of RPA was to utilise a smartphone as the source for demonstration content. Through research and analysis of both observational learning and the target hardware, the RPA was constructed by an interdisciplinary development team. When designing any technology to support education and training, the usability of the hardware and software is as essential as the content. The system's usability describes how effectively and efficiently the desired goals can be attained in a specified context of use and the user perception of this process (ISO, 2018). This definition values both objective performance and perceived achievement. Poor usability will deter the provider and receiver of information in an educational setting. If students cannot quickly learn how to use a tool, the instructor will need to spend more time educating and troubleshooting the tool (Akçayır & Akçayır, 2017). Outside of a structured lesson, students may avoid the technology entirely as they do not have the technical support of the instructor. In an immersive virtual environment, technical frustrations or usability issues can break the psychological sense of presence. The user objectively sees and subjectively feels that they are not in the real world but in the synthetic virtual setting (Slater & Sanchez-Vives, 2016). Breaking presence will engage the user with the real world and distract them from the virtual experience. To evaluate the direction of applications in development, regular testing is crucial. Features developed from informed design still need to be tested to see if they are fit for purpose and usable by the target audience within a given context. Lessons learnt early can inform future iterations of development. A usability study was conducted to evaluate the functional prototype RPA and inform the ongoing design process of the application.

This study aimed to:

1. Measure the memory recall of movements observed.
2. Evaluate usability, ease of use, and perceived learning effectiveness.
3. To compare RPA across different platforms.
4. Establish guidelines for developing virtual demonstration-based content and delivery.

2.4 Design of the Recovery Position Application

The RPA displays two virtual avatars performing the recovery position sequence. The avatar performing the sequence is named "Helper". The avatar placed into the recovery position is named "Casualty". The design of RPA was based on the need for observational learning within a DBT framework and the hardware considerations for mobile VR.

The two areas of focus to inform the design process were:

1. Technical limitations of the hardware and Google Cardboard platform.

2. Requirements of successful observational learning of a demonstration.

2.4.1 Hardware Considerations of Mobile Virtual Reality

Mixed reality describes a broad set of technologies that combine real and virtual worlds for interactive experiences. The spectrum of platforms ranges from entirely synthetic, virtual environments to authentic environments (Milgram & Kishino, 1994). At one extremity of this spectrum is virtual reality (VR). VR is a computer-generated synthetic world that responds realistically to human senses and thus creates the illusion that the user is in a new reality (Slater, 2014). The term "Immersive Display" describes how the virtual world surrounds the user's visual sense with replication of a stereoscopic view (a display for each eye) and human perspective (closer objects appear bigger) (Slater & Sanchez-Vives, 2016). An example of an immersive display would be a head-mounted display (HMD) or wall projection system like the CAVE (Cave Automatic Virtual Environment) (Figure.2.1). In such displays, the position and rotation of the user's head are tracked in 3D space and updates the display accordingly. This positional and rotational tracking is described as six degrees of freedom (6-DOF). This relates to the three axial points (x, y, z) recorded for both position and rotation. As the user turns or moves their head, it appears they are turning their head within the virtual world.

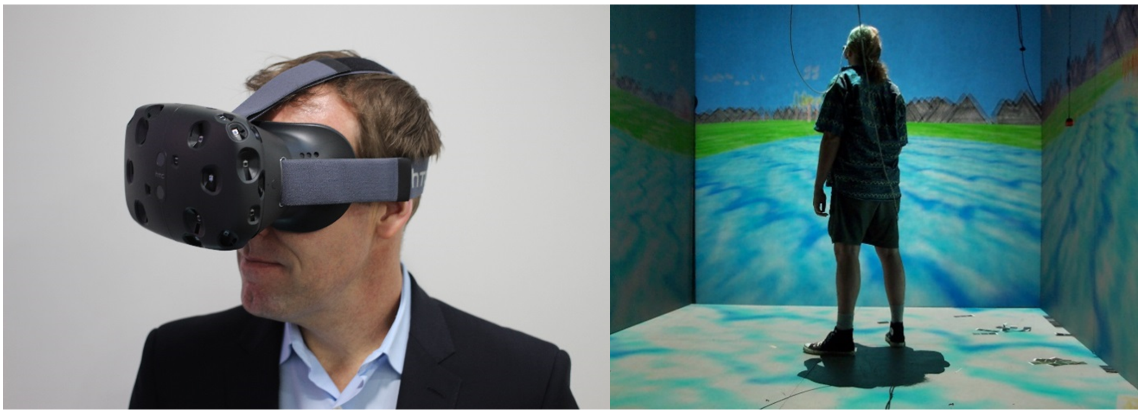

Figure 2.1: Virtual Reality Examples: (Left) HTC Vive headset that is tracked in 3D space. (Right) A CAVE where the displayed image is projected onto a surrounding surface from the user's perspective.

A mobile phone can deliver a similar mobile VR experience. A mobile phone is placed inside a headset close to the user's eyes to become the display (Figure.2.3). Although a mobile phone display will have much less graphical power than a PC, it will need to render acceptable framerates at clear resolutions to minimise visual lag of movements.

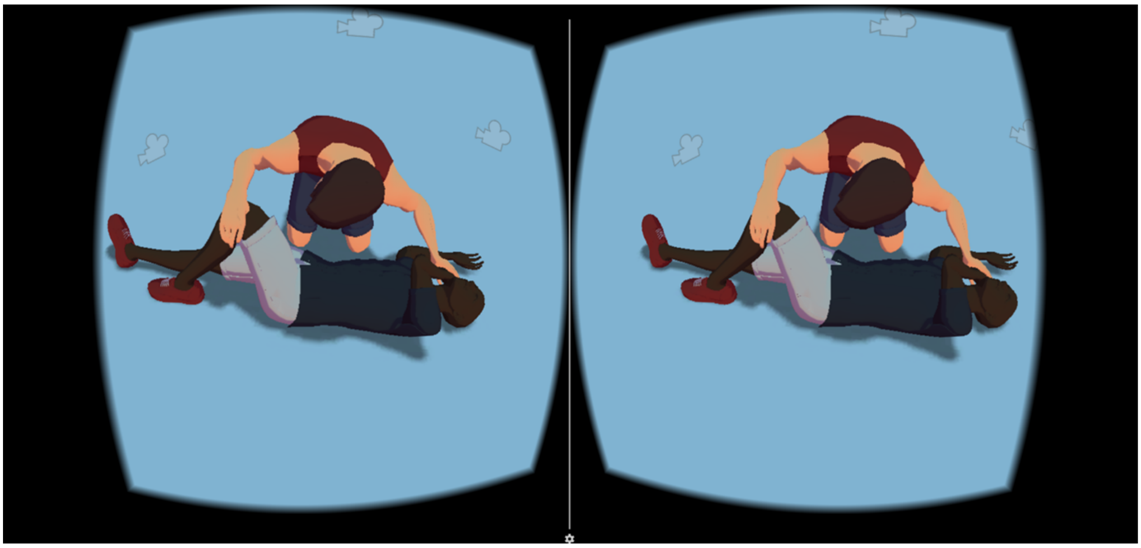

Figure 2.2: Shows how immersive displays render to each eye on the same display creating stereoscopic vision. Graphics of the application are informative but optimised.

Visual lag can cause simulation sickness (Davis, Nesbitt, & Nalivaiko, 2014). This does raise a design consideration that graphics should be informative but need to be optimised stylistically to maintain graphical performance (Figure.2.2).

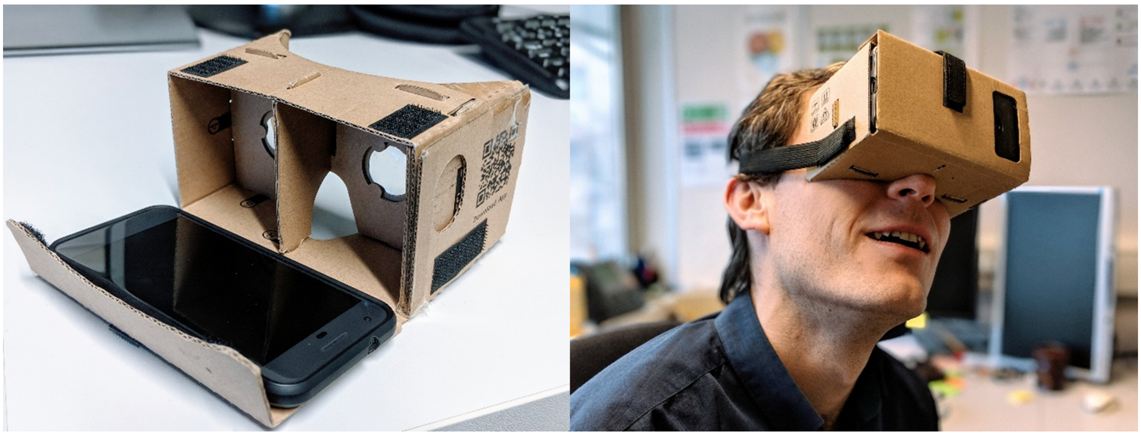

Figure 2.3: (Left) Shows how a mobile phone is placed as the display for Google Cardboard head-mounted display. (Right) Shows the placement of the Google Cardboard display in use.

Selection techniques refer to how a user interacts with the graphical user interface of an application to make choices of how to progress and change settings. Although some dedicated mobile VR hardware configurations will have a connected controller or a single HMD button to press, this is not the standard. To increase the accessibility of the RPA (Both dependent on technology and the potential disability of a student), the assumption will be that users will have no more than a smartphone and an HMD housing. For this reason, a time-on-target selection method was used. This method allows all selection choices to be controlled with the movement of the user's head. A black circle in the centre of view acts as the user's reticule (which represents the relative centre of the display). When the user rotates their head and positions the reticule over selectable icons, it will enlarge, disappear and then slowly draw a circle. Once the circle is complete, the icon is selected. The delay caused by the circle draw facilitates an intended action and reduces the chance of accidental selection (Figure.2.4).

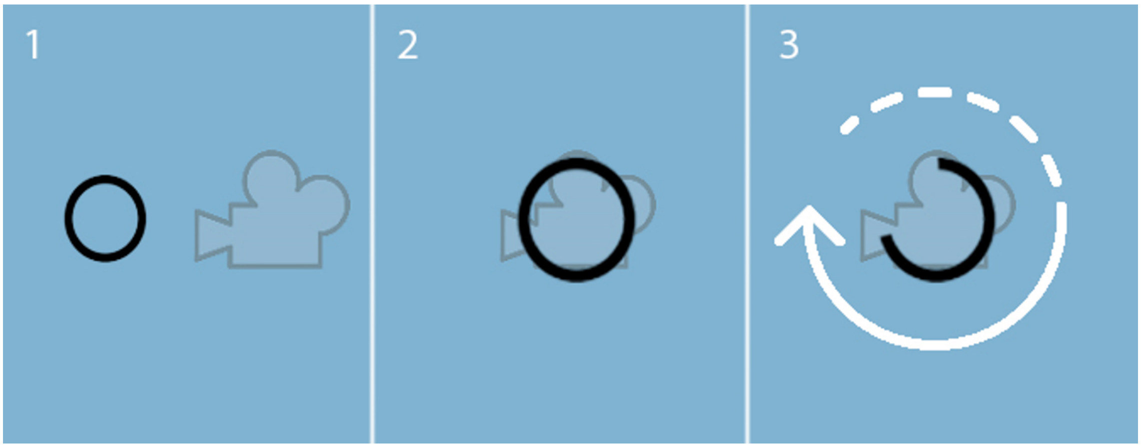

Figure 2.4: Image sequence showing time on target selection method of camera icon (panel 1). The user reticule is positioned centrally to the viewport (Black circle, panel 1). The user positions the reticule over the camera icon (panel 2). Reticule expands to provide feedback that the icon is selectable. The reticule then redraws itself over 2 seconds (panel 3, white arrow shows the direction of redraw). When the circle is fully redrawn, the item is selected.

A comparative limitation of mobile VR is that only rotational information is tracked in 3D space (through the phone's accelerometer). Positional information is not. This is known as three degrees of freedom (3-DOF). In practical terms, the user can rotate their head, but there will be no visual feedback of translational movement (walking, crouching etc.). Without the visual update, the translational movement could lead to simulation sickness and break the sense of presence. Due to this limitation, the user should be seated when using mobile VR. A teleportation system was developed to enable navigation of the 3D virtual environment without a controller input or tracking of translational movement. Camera icons visualise teleportation locations. Once a camera icon is selected, the user will teleport to that location.

2.4.2 Delivery of Observational Content

This initial iteration of the RPA delivers a virtual demonstration to create a mental representation of the action sequence. To inform design and implementation, we used an observational learning approach. Key elements and considerations are described below.

Cognitive load theory describes how the processing of the learning task (intrinsic), task presentation (extraneous), and the mental resources devoted to assimilating information into long-term memory (germane) utilise an amount of information within working memory (Sweller, 1988; Van Merrienboer & Sweller, 2005). As cognitive load increases, fewer mental resources are available to explore learning scenarios and assimilate information. Showing all concept elements can be too much new information for a student to interact with and apply (Pollock, Chandler, & Sweller, 2002). This suggests that breaking down a movement sequence into individual actions may reduce the intrinsic cognitive load (Yang, Leung, Yue, & Deng, 2013). When using technologies or learning materials unknown to the user, there will be an increase in extraneous cognitive load. Extraneous cognitive load can also be increased by overloading sensory streams. For example, if all information is presented visually (writing, diagrams, animation), then the visual stream can be overloaded. Directing some information through the auditory stream (written text to verbal oration) reduces extraneous cognitive load and aids learning. This is known as the dual modality principle (Low & Sweller, 2005). The RPA uses narration to accompany the visual demonstration that describes actions being performed by the avatars.

When observing actions, students perceive the spatial coordinates of an instructor's movement in relation to the demonstrator's body. This provides a reference for body position and speed of action. To imitate these actions, the student must mentally map this information to their own body. This requires transcoding information from an allocentric spatial frame (objects are located relative to one another) to an egocentric spatial frame (objects are located relative to the learner's body) (Willingham, 1998). The transcoding will require an amount of cognitive load based on the learner's ability to rotate mentally and familiarity with the action (Krause & Kobow, 2013). This increase in cognitive effort is evidenced in motor skill imitation studies. Participants took less time to imitate hand actions when viewed from an egocentric spatial frame, compared to an allocentric one (Jackson, Meltzoff, & Decety, 2006). Physically aligning the demonstrator's allocentric spatial frame with the student's egocentric spatial frame will reduce the extraneous cognitive load (Krause & Kobow, 2013). Therefore, the observational content within a virtual environment should allow multiple vantage points that enable user navigation between allocentric and egocentric perspectives, as in our implementation.

Learner autonomy over navigation around an object (Brooks, 1999) has been shown to improve the memory recall of complex 3D objects and spatial layouts. Participants who memorised the layout of a virtual building recalled more when in navigation control. Similarly, when participants rotated a virtual inner ear model (Jang, Vitale, Jyung, & Black, 2017), those with control over the direction of rotation could draw this anatomical structure more fully. This prior research established that autonomy over navigation and the flow of an experience could aid the spatial and episodic memory of what is observed. Virtual observational content should therefore give control to the learner over how they travel, their choice of perspective, and the pace at which they explore the observed information. In our implementation, users can teleport between observational viewpoints and control when the following sequence is enacted (Figure.2.5).

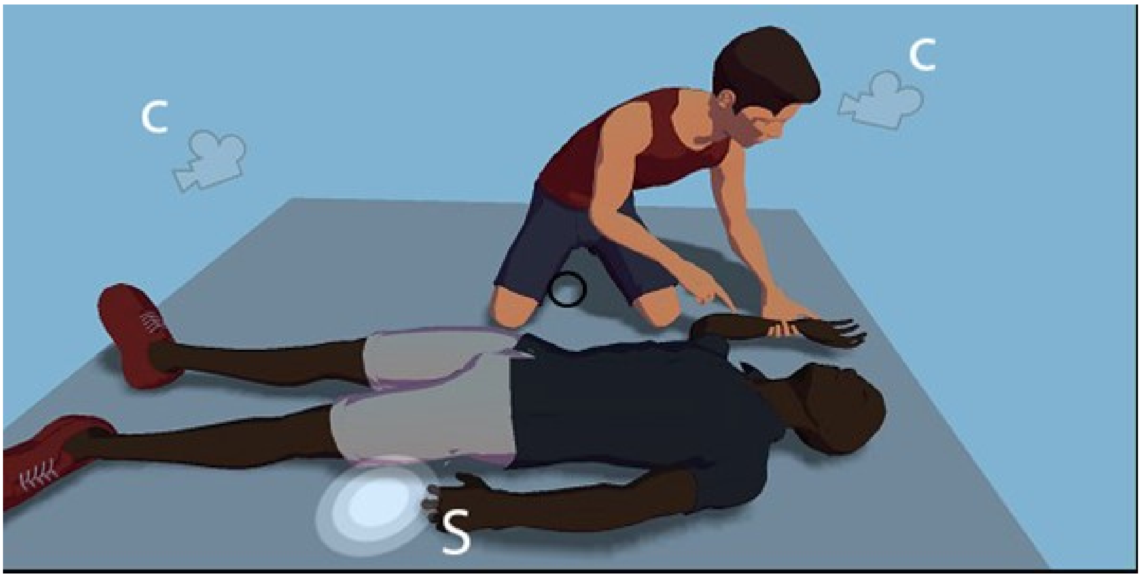

Figure 2.5: Users can teleport to the camera locations (Marked "C") at any point during the demonstration. The demonstration will not progress to the next movement until users select the "Next Step" icon symbolised by a white circle (Marked "S").

When observing actions for later memory retrieval, recognition or imitation, the movement must be demonstrated accurately in terms of body posture and time spent transitioning between poses. From brief observation of a body posture, and with no attempt to imitate, we can accurately remember and recognise action poses (Urgolites & Wood, 2013). Action can be understood even when abstracted into 2D images or when the action is described verbally. However, motor skill acquisition is more effectively taught through animations than still pictures (Höffler & Leutner, 2007). As long as the focus of the subject matter (in this case, analysis of human movement) is represented accurately, then there is little benefit in raising the fidelity of graphics (Norman, Dore, & Grierson, 2012).

Neurological studies of observing actions suggest similar neurons fire when an action is performed and when passively observed (Rizzolatti & Sinigaglia, 2010). Through this mirror mechanism, it is suggested that we internally simulate performing an observed action to predict possible action. Our understanding of the observed action mediates this neurological representation. For example, Iacoboni et al. (2005) showed participants images of grasping a mug with no context and within the context of breakfast. FMRI (Functional Magnetic Resonance Imaging) scans showed more potent activation of mirror neurons when the context was established. The observer's own goals can influence the pattern of mirror neurons that discharge. In previous work by Molenberghs et al. (2012), participants were asked to observe hand actions under three different mindsets: Understand the meaning of an action, observe physical features, or passively view the actions. FMRI scans showed subtle variations in mirror neuron discharge depending on this mindset. These two studies show that higher-level cognitive processes (mindset and context of a situation) mediate the neurological representation of movement when observing actions. Suppose participants are told that they are playing a computer game to improve their golf putting ability instead of simply enjoying the experience. In that case, they will show better real-world improvement (Fery & Ponserre, 2001). Establishing the context of the learning activity and observed demonstration motivates the user to learn and aids the cognitive representation of actions. The RPA establishes context by combining an instructional voice-over and text overlay that informs the user of the application's educational purpose and goals.

To summarise, immersive displays facilitate an egocentric perspective within a 3D virtual environment and replicate movement observations as if in the real world. The mental representations of these movements may aid the understanding of actions throughout motor skill development. The core requirements of effective observational learning of action sequences via demonstration are: breaking down a movement sequence into manageable chunks, spatial representation of actions, manipulation of spatial frames of reference, user control, accurate representation and context.

2.5 Methodology

2.5.1 Procedure

This study received approval from the Research Ethics Integrity Committee, University of Plymouth. 61 (female = 6) participants were recruited for this study. This participant pool was a convenience sample of students studying game development at the University of Plymouth (N = 57) and industry professionals who develop games (N = 4). Participants were assigned to one of three groups. Group one (N=20, 2 female), titled "Mobile VR", used mobile VR to interact with the application. Group two (N=20, 1 female), titled "Desktop Display", used a desktop PC display with a mouse and keyboard for navigation. Group 3 (N=21, 3 female), titled "Video", watched a video recording of the RPA on a desktop PC display. The use of the Desktop Display group was to test the difference in recall between an immersive mobile VR and a non-immersive display. The Video group was used to explore the role of autonomy on memory recall compared to the Desktop Display group. By watching a video of the RPA, the participants would be viewing the same content without controlling the flow of information and choice of perspective.

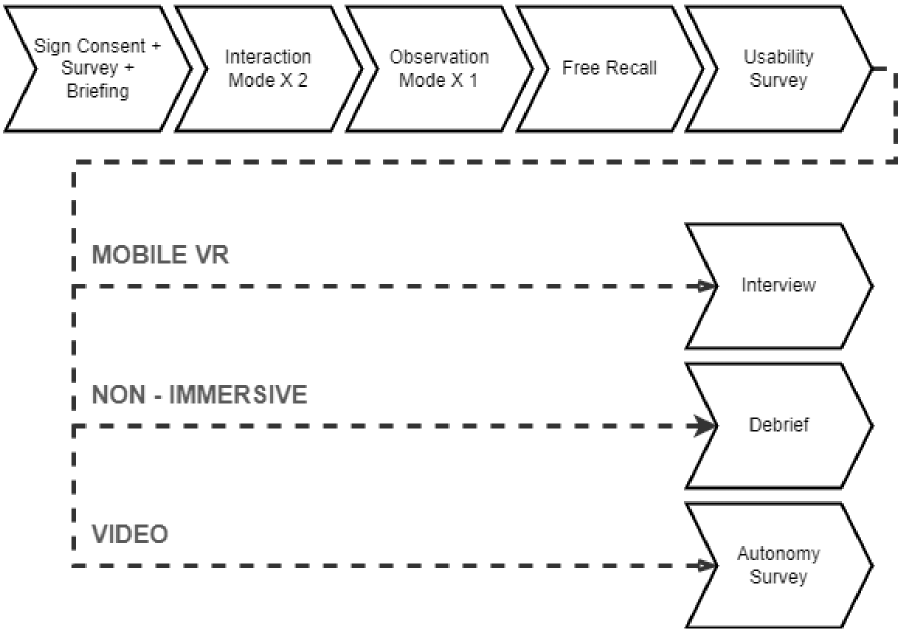

Figure 2.6: Flow diagram to illustrate the sequence of events encountered by participants in this study

The same procedure was used for all groups (Figure.2.6): Participants filled in a questionnaire that extracted basic demographic information, including one question to assess their perception of recovery position knowledge. This question asked, "How would you describe your knowledge of the recovery position?" with the option of none, poor, intermediate, or advanced as potential responses.

For the Mobile VR and Desktop Display groups, verbal instructions on using the RPA were then given. Participants were seated and asked to play the "Interaction" mode twice. During interaction mode, participants selected when to move between the individual steps of the recovery position. Participants could teleport between viewpoints to change their viewing angle. A verbal description accompanied each step of the visual demonstration. Participants were then invited to review the "Observation" mode with the avatars enacting the recovery position action steps in sequence automatically. No verbal description was present in this mode. This process took participants 15 minutes to complete. The Video group watched a recording of the RPA through the above steps.

Post-exposure to the RPA, we tested the participant's memory recall of the recovery position sequence. Participants could write down what they remembered or verbally report it to the investigator. Participants were then surveyed, exploring the usability of the application and perceived usefulness as an educational tool. For the Mobile VR group, the participants were then interviewed about their experience using the Recovery Position Application. The memory recall test and the interview were digitally recorded with the participants' permission.

When surveyed on usability, the Systems Usability Scale (SUS) survey was administered (Brooke, 1996). Usability describes how satisfying and effective it is for users to achieve their goals whilst using a system (ISO, 2018). With good usability, users will be focused on the task they need to complete and not distracted by the frustrations of using a tool. Poor usability can therefore interfere with software goals and in an educational context, learning goals. The SUS gives an overall score out of 100. Numerical bands within this score can be abstracted to more understandable adjectives (good, poor, excellent etc.) (Bangor, Kortum, & Miller, 2009). Additionally, the survey questions can be categorised into "how usable" and "how learnable" the system is (Lewis & Sauro, 2009). The SUS survey can help highlight issues and current successes that inform future development. An additional question was added to the SUS survey: "Would you use this application as a learning resource?" to explore participant perceptions on whether they see value in using the system as an educational tool.

After the usability survey, eight of the Mobile VR group were invited to participate in a semi-structured interview (10-15 minutes). The goal of this interview was to help inform the development of RPA by establishing feature set refinements that would improve usability and learning and catalogue any bugs encountered. To achieve this, participants were encouraged to reflect upon their ability to navigate RPA, engage with the content, and to what extent RPA has added to their knowledge of the recovery position. Audio from interviews was recorded with the permission of the participants. These interviews were then summarised during data processing to highlight feature requests, bugs, and thoughts on usability and learning.

2.5.2 Data Analysis

RPA demonstrates movements and highlights key assessments to be made by an individual before they perform a recovery position action. The RPA delivers 27 details about the recovery position (See Table.2.1). To deliver this content, RPA uses narration and avatar demonstration where appropriate. The use of narration in the context of visuals (dual modality) can aid working memory resources and can be viewed as good practice when viewing a dynamic visualisation (Low & Sweller, 2005). However, some assessment details described in RPA do not have a dynamic visual counterpart. For example, when instructed "Check the area poses no risk to yourself", the avatars are kept in a static pose awaiting the start of the first recovery position movement. Therefore, the visuals do not demonstrate the act of checking the area for risks. This may reduce visual memory hooks and context for this narrated information and thus reduce recall of the assessment details. Therefore, on analysis, details that have audio (narrative) representation only, have been highlighted to examine recall concerning details that have both visual demonstration and audio.

Information delivered through the Recovery Position Application

Step No.	Detail No. Per Step	Detail Description	Modality Delivery
1	1	Check the area poses no risk to yourself	Audio
	2	Check that the casualty is breathing	Audio
	3	Gently tilting the head back	Audio + Visual
	4	Listen and feel for breath on your cheek	Audio + Visual
	5	Look for movement in the chest	Audio
	6	Only proceed if they are breathing normally	Audio
2	7	Select arm closest to you	Audio + Visual
	8	Place at right angle to casualty's body	Audio + Visual
	9	Palm facing up	Audio + Visual
3	10	Select hand furthest from you	Audio + Visual
	11	Bring across casualty's body	Audio + Visual
	12	Place back of casualty's hand against patient's cheek	Audio + Visual
4	13	Grab knee furthest from you	Audio + Visual
	14	Raise it up	Audio + Visual
	15	Until foot is flat on the floor	Audio + Visual
5	16	Gently roll casualty towards you	Audio + Visual
	17	By pulling on the knee	Audio + Visual
	18	Support the casualty's head with your hand during this manoeuvre	Audio + Visual
6	19	Tilt the head	Audio + Visual
	20	By lifting the chin	Audio + Visual
	21	Ensure airway is open	Audio + Visual
	22	Check for normal breathing	Audio + Visual
7	23	Select top leg	Audio + Visual
	24	Bring out at right angle	Audio + Visual
	25	To support the casualty	Audio + Visual
8	26	Call an ambulance	Audio
	27	Monitor the casualty	Audio

Table 2.1: Table to show material delivered through the recovery position application. The steps indicate when the user can select to move on with the demonstration. The details per step describe the individual information elements within each step and represent the amount of material delivered at each segment of the sequence. The detail description synthesises the audio and visual information delivered for each detail. The modality delivery indicates if this information uses audio and visual representation, or audio only to deliver each detail.

2.6 Results

2.6.1 Participants

Most participants were aged 18 – 24 (N = 53). Four participants were aged 25 – 34, two aged 35 – 44, and two aged above 44. Out of the 61 participants, the majority (N = 57) felt as though they were "technologically savvy", and most had used some form of immersive virtual technology before (N = 47). When surveyed on their perceived knowledge of the recovery position before exposure to the raining applications, the majority (32) perceived their knowledge to be poor. (See Figure.2.7.

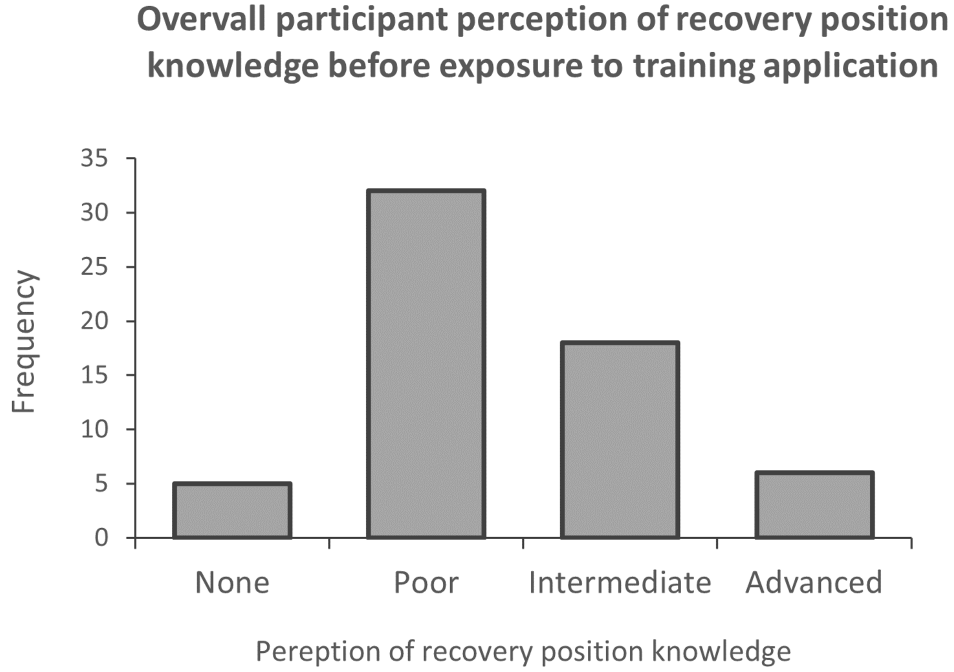

Figure 2.7: Chart to highlight participant response to "How would you describe your knowledge of the recovery position?" before using the training application. The majority (32) perceived their knowledge to be poor.

A three-way ANOVA was performed to investigate if previous knowledge of recovery position, experience with immersive VR technology, and study group (Mobile VR, Desktop Display, Video) affected the recall test.

The three-way ANOVA did not observe any statistically significant interactions between these groups nor individually as a main effect. A one-way ANOVA was performed to compare if the perceived previous knowledge affected recall accuracy across all study groups. The one-way ANOVA did not observe a statistically significant effect between knowledge perception groups and final recall (F(3, 57) = 0.138, p = .937).

2.6.2 Recall

Overall recall accuracy scores (1 = 100% accuracy) were calculated for Mobile VR, (M = 0.69, SD = 0.17) Desktop Display(M = 0.62, SD = 0.51), and Video (M = 0.70, SD = 0.13) groups. To assess if there was any significant difference between the recall scores and platforms used to display RPA, a one-way ANOVA was conducted between cohort groups (Mobile VR, Desktop Display, and Video) on recall. This test did not observe any significant difference in recall between cohort groups (F(2, 58) = 2.648, p = .079)

Recall of details delivered through both audio and visual modalities (narrative as well as avatar demonstration) was compared to using audio modality alone (narrative description alone), see Figure.2.8. A paired sample t-test was administered to see if there was an effect of delivery modality on recall accuracy. The results suggest a significant increase in recall using audio with visual (M = 0.671, SD = 0.152) modalities rather than audio modality alone (M = 0.470, SD = 0.239); (t(60) = 7.949, p = <.001).

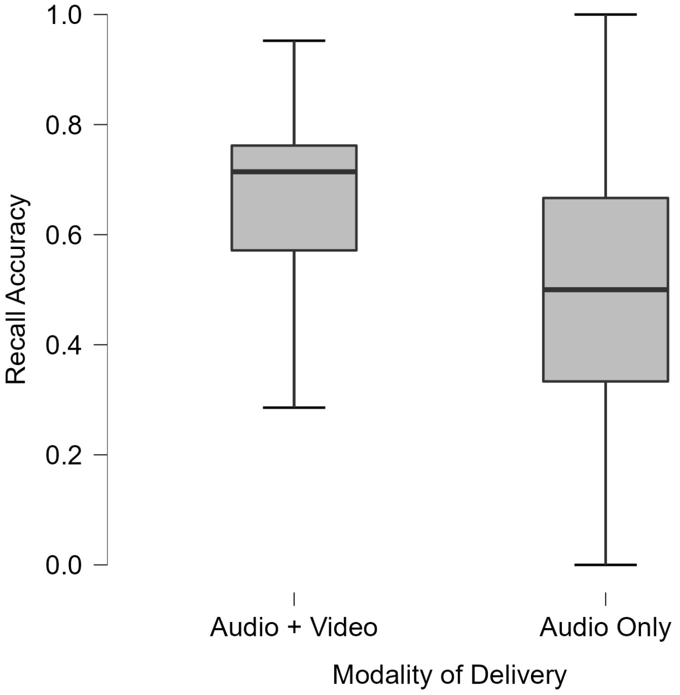

Figure 2.8: Box plot comparing recall of details using either visual and audio information or audio information alone. The median point of both audio and visual delivery is above the interquartile range of the audio-only boxplot. This suggests a significant difference between these groups as confirmed by a paired sample t-test

To assess if there is significant variation in recall across the sequence of steps within RPA, a repeated measures ANOVA was conducted using recall for each step as a level. Recall across the steps was statistically different (F(2,58) = 32.249, p = <.001). Figure.2.9 highlights that steps 1, 6, and 8 had the lowest accuracy score, with under 50% of information recall. Steps 2, 3, 4 and 5 have a 69% -93% range in the average recall. Ignoring step 1, the overall trend in the recall is that initial information is recalled best, with a steady decline in accuracy as the steps progress.

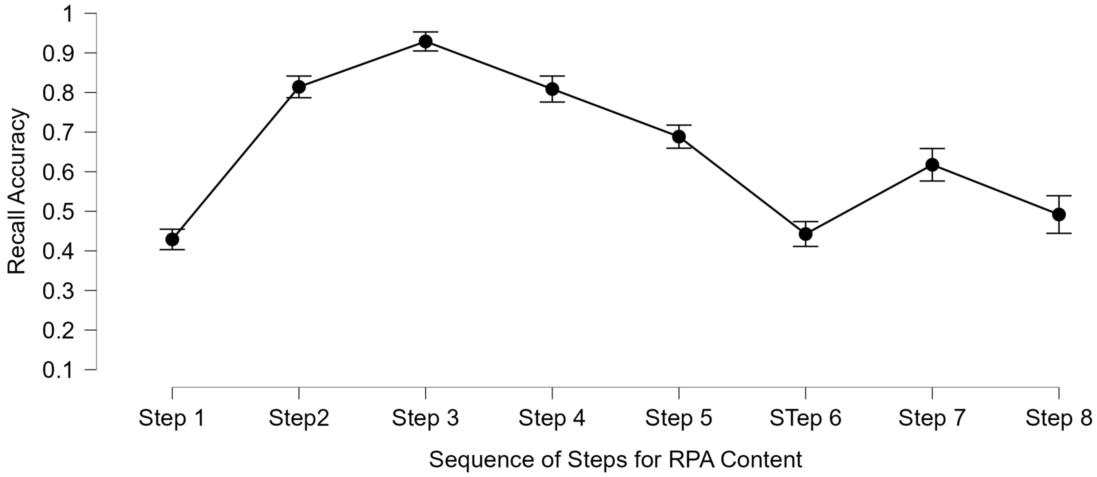

Figure 2.9: Graph to show recall for each step in the sequence of information delivered to users of the RPA. Error bars highlight significant differences between steps. Of interest, steps 1, 6, and 8 have the lowest recall. Generally, there appears to be a decrease in recall as the steps progress.

2.6.3 Usability

The overall SUS score for the Mobile VR group was 82.50 (SD = 11.09), which can be described as "good". The SUS score for the Desktop Display group was 87.63 (SD = 8.01), which can be described as "excellent". The SUS score for the Video group was 80.83 (SD = 10.56), which can be described as "good".

To assess if participants had positive or negative perceptions of how usable and how learnable the system was, the responses to survey questions associated with each category (usable and learnable) were averaged. This average (1-5 form the Likert scale) was compared against the mid-point Likert scale value of three. If the resulting value fell significantly below three, this would be viewed as a negative perception and above as positive. Table.2.2 shows that participants in all three groups (Mobile VR, Desktop Display, and Video) perceived the Recovery Position application as both usable and learnable. In response to being asked if they would use this application as a learning resource, participants were significantly positive to this statement across all groups.

A one-way ANOVA was conducted to assess if there was any significant difference between groups for SUS scores and perceptions of use as a learning resource. No significant effect for these factors was observed: overall SUS score (F(2, 58) = 2.551, p = .087), "how usable" scores (F(2, 58) = 2.866, p = .065), "how learnable" scores (F(2, 58) = 0.670, p = .516), and use as a learning resource (F(2, 58) = 2.851, p = .066).

Usability Survey Results

Usability Category	Mid-Point	Mean ± SD	df	t-value	p	SUS Score / Descriptive
Mobile VR						
How usable	3.00	4.26± 0.41	19	13.79	< .001	82.50 / Good
How learn-able	3.00	4.45± 0.79	19	8.18	< .001	
Learning resource	3.00	4.40± 0.60	19	10.47	< .001	
Desktop Display						
How usable	3.00	4.46± 0.35	19	18.86	< .001	87.63 / Excellent
How learn-able	3.00	4.68± 0.47	19	16.05	< .001	
Learning resource	3.00	4.75± 0.64	29	12.25	< .001	
Video						
How usable	3.00	4.16± 0.48	20	11.05	< .001	80.83 / Good
How learn-able	3.00	4.55± 0.55	20	13.00	< .001	
Learning resource	3.00	4.24± 0.83	29	6.83	< .001	

Table 2.2: SUS scores are used to categorise how useable and learnable the software is. P-values from the SUS scores are significantly above the mid-point of 3.00 from the Likert scale. This suggests a positive response to the usability questions. The overall SUS score is given with an appropriate adjective description relating to that score.

2.6.4 Autonomy

Group three (Video) were asked a further four questions in the post-exposure questionnaire to help evaluate usability when not having control over navigation and flow of content.

- Q1 - Would you prefer control of navigation and flow of information?
- Q2 – Why is this?
- Q3 - Did you find the lack of control in navigation and flow of information frustrating?
- Q4 – Why is this?

18/21 responded "yes" to Q1. 9/21 of the participants responded "yes" to Q3. This suggests that participants would prefer control of the flow of information, but 12/21 did not find the lack of control frustrating. In response to Q2, 11/21 participants valued: "learning at my own pace", 4/21 valued being able to repeat a step when needed, and 3/21 valued control over the camera viewpoint. The lack of these elements was cited as the cause for frustration when answering Q4. However, those that were not frustrated, explained that information was delivered at a suitable pace.

2.6.5 Interviews

Most of those interviewed from the Mobile VR group described navigation as "fairly easy" to learn. Most participants required further clarification before they were confident in using RPA. For example, "the navigation needed explaining to me", and "I am not sure I would have seen the cameras if not told about them".

However, once they had this guidance, they felt the system was easy to use. "Quickly got to grips with it. . . .. Moving the camera was fine", and "Very simple control methods". Participants highlighted improvements for general interactions. In particular, the selection method was not as quick or apparent as all participants would have liked. "Only issue I found was time on targets takes significant time, and you can miss the animation during this time", "..there were some moments were hard to see the target to focus on".

One participant also found the hardware challenging to adjust "Found headset awkward and focusing was not easy. If there was some text that could be used to as a point of reference for focusing that would be useful."

Participants reported that they predominantly only used two view positions. These were the front and back, as they facilitated oversight of the entire action sequence. A common request was to add two more camera positions. One directly above the demonstration for an overview. One directly from the viewpoint of the "Helper" avatar. The latter viewpoint was to experience observation of the action sequence as if performing the recovery position. Viewpoints positioned near the head of the casualty were deemed too close, as the helper's body obscured details of the movement. The proximity of these viewpoints to the casualty forced users to translate their head position for a better vantage. This highlights a limitation of mobile VR. The positional translation of the HMD is not tracked in 3D space and limits the user's natural head movement to rotations only.

Participants described this limitation as frustrating.

Participants viewed the graphics as believable, even though they did not describe the style as realistic. A key driver for this believability was the perceived realism of the animations. One participant elaborated on how they expected the graphics not to be realistic: "It might be cartoony, but that is what you expect from an app. You don't expect a real-life person".

Participants felt they could clearly see the recovery position demonstration "Yeah, for the most part it was easy to follow, slow enough animations.". Additionally, the different camera positions aided a complete understanding of the sequence, "Occasionally, I had to change cameras to double-check movements". However, there was some difficulty in registering smaller movements from the demonstrating avatars, "I found the hand on the cheek slightly harder to see" and "I would like More emphasis on smaller details like head movement.".

The majority interviewed felt as though they had learnt much about the recovery position, "Was a good refresher of previous knowledge", with one feeling as though they could now practice the sequence "I feel I would be happy to perform the recovery position". One participant felt that it was more novelty than useful emphasising, "I don't feel I would learn much from this."

Both audio description and visual demonstration was perceived as an effective combination for information delivery. Some participants admitted that they only listened to the descriptions as they found this easier to assimilate. Some ignored the audio description completely. The majority found helpful information in both. Participants also noted that they did not mention the first detail (check that the area is safe) because there was no demonstration of this action, just audio. Secondly, there was no visible danger in the

	Feature Development
1	Starting room to adjust headset and check focus is clear
2	Cameras can be reduced as front and back are most useful. One above and one directly from the rescuer's perspective
3	More environment detail for context and ability to assess if it is safe or not
4	More context on when to and when not to perform the recovery position
5	Repeat animation function
6	Pause animation function
7	Emphasis on smaller details of the sequence, for example, head movements
8	Add extra steps for smaller movements for clarity

Table 2.3: Table to highlight feature improvements and additions from interview discussions after using the Recovery Position Application on a mobile VR platform.

environment.

Discussions from interviews highlighted feature additions and adaptations that the development team could explore to aid the usability of RPA and learning of the recovery position content. These are summarised in Table.2.3

2.7 Discussion

This study examines the recall from and the design implementation of RPA, an application that uses 3D avatars with narration, to demonstrate the recovery position sequence. Three groups were used to compare the memory of RPA content when observed on an interactive mobile VR platform, an interactive desktop display platform, and a non-interactive video. The design of RPA was evaluated through usability analysis and participant discussion to inform future development.

Across all three platforms, participants could recall most of the correct body positions for each action and the correct sequence order, despite using the RPA for a short duration (10 minutes). This suggests that a cognitive representation of the recovery position was developed effectively across platforms.

There was no significant difference in recall between the three groups. This suggests that RPA could be used on multiple mixed-reality platforms with a similar effect on demonstration recall. Within the DBT framework, demonstrations of action sequences could be supported by many display devices. The displays used depend on the needs of the training or the resources at hand. Of interest is that a more immersive device (Mobile VR) did not aid recall of action sequences compared to a non-immersive display (Desktop Display). This may suggest that a desktop display provides enough spatial information and effective egocentric/allocentric perspectives to create a mental representation of observed movements. In addition, the limitations of mobile VR platform translational movements not tracked and a small field of view may weaken any benefits a dedicated immersive VR (HTC Vive, CAVE) set-up may facilitate observational learning of demonstrations.

This study used a convenience sample primarily of those studying game development and some industry game developers. This sample has similar age and technological understanding characteristics. A wider sample is required for these attributes to be drawn into the analysis of their impact on recall. Previous perceived knowledge of the recovery position was varied in this sample and did not statistically impact recall. Perhaps even when participants felt they had some good knowledge of the recovery position, it has been a long time since they have had to put it into practice. Therefore individuals had a good idea of what was required or the key goals of the recovery position but needed help to articulate the details of these steps before exposure to RPA. Therefore those that felt they had poor or good knowledge of the recovery position had a similar understanding of the individual movements. The method for collecting previous recovery position knowledge was based on participants' perceptions. This could be improved with an additional knowledge test to evaluate actual knowledge and not just subjective perceptions. Since most movement details were recalled at all levels of perceived previous knowledge, there is good potential for RPA to establish and maintain a mental representation of the recovery position movements across knowledge levels.

Looking at the recall across steps, there is a general trend of recalling more from the beginning of the RPA that reduces steadily across steps. This is similar to the primacy effect from list recall, where the initial few words are robustly recalled at the beginning of a list (Murdock Jr, 1962). The rationale for this phenomenon is that individuals have more time to consolidate the information they first encounter and, therefore, can put it into longer-term memory more efficiently. This would account for the reduction of information recall across the steps. However, step one had the lowest recall. This step was made mainly of audio-only narration and may account for the relatively poor recall. With visual representation, narration may create stronger memory hooks or emphasis, compared to visual and audio stimuli. Additionally, when the audio narration was used on its own, the information was assessment-based and not describing movements. It is possible that the participants found it easier to mentally represent movements as opposed to judgements. Another reason for poor step one recall was that the focus of RPA was to develop an action sequence understanding, and the participants were informed of this before using RPA. A possible result could be that participants focused more on movement information due to this explanation. These judgements would have more context when applying the recovery position but may not be seen as useful information until the movement has been learnt. This would again bias participants to focus on the movements before judgements. From a working memory perspective, there may have been too much information in step one. Step one had the lowest recall rate and the highest amount of details. This could indicate that information in this step exceeds limited working memory (Miller, 1956). When working memory is exceeded, the individual will either ignore any extra information or may develop a method to organise it into smaller chunks (Yang, Leung, Yue, & Deng, 2013). The combination of audio-only, focus on an action sequence, and a large amount of information suggests that step one is at a disadvantage compared to other steps. Due to the primacy effect, if the initial step had equal emphasis, and was segmented into smaller chunks, this step should be remembered more regularly than other steps.

When observing the RPA demonstration, actions performing head adjustments to the casualty were least recalled, as shown with a lower recall rate for step six in the sequence. This can be explained in terms of the learner's goals. In goal-directed imitation theory (GOADI) (Wohlschläger, Gattis, & Bekkering, 2003), the goal and intent of the movement supersede the act of imitation. For example, in flicking a light switch on, the individual is not concerned with how this is achieved (correct arm direction, which part of the hand to use etc.) but focused on the goal of the action (moving the switch up). Similarly, in this study, as participants acquire a mental representation of moving a body into the recovery position, they focus on the general goal of a body position. Although important, the smaller details of this act may not be the focus of the participant's goal. In this case, placing a casualty into the end body position of the recovery position is the user's primary goal. Feedback from interviews also suggests that the movement of the helper avatar's hands was not clear when viewed from an obstructed angle. Separating head and hand actions into a single step, with more detail, could aid the recall of these movements.

The video condition removed participant autonomy over navigation and flow of information. Survey results and interviews frequently suggested that users would like more autonomy when exploring the RPA. However, removing autonomy did not translate to poorer memory recall. A more complicated movement set or longer exposure to the RPA in this format may reduce memory recall due to a lack of autonomy. However, for short (10-minute) demonstrations, this study suggests that the autonomy of information has little impact on memory recall and a minor negative effect on usability.

The usability survey observed positive perception for the RPA across Mobile VR, Desktop Display and Video groups. At the same time, the interviews enabled exploration of where participants felt confident and had difficulties with RPA. According to the overall SUS score, Mobile VR and video viewing groups have good usability, whilst the Desktop Display group averaged an excellent usability standard. All three platforms elicited significantly positive perceptions for being learnable and useful in the context of a learning tool. This was reflected in interviews where many felt able to use RPA once they understood the controls. This set of results points to the interaction design of RPA being sufficient for novices to pick up and understand relatively quickly across platforms. Interviews highlighted feature additions and adaptations to improve usability and engagement with the recovery position content. These discussions suggest emphasising more minor details of the recovery position through smaller steps and visual highlighting to guide attention. Additionally, a repeat and pause step functionality would help encode details missed through a step. Both these additions help the recollection of smaller movements. Some further development was identified with camera placement and the need for a tutorial. Participants mentioned they felt they had learnt a lot about the recovery position sequence and, when surveyed, felt that they would use such a tool for training. This suggests that the implementation of interactions allowed sufficient engagement within the RPA content. However, this can only be claimed in relation to the convenience sample of game development students and industry professionals. This type of sample may have more experience and intrinsic motivation to use virtual content for training across various technological platforms.

Between groups, there was no statistically significant difference between usability scores. This suggests that when exposure to the RPA is short, it can be easily interacted with across multiple platforms, even though it was designed around mobile VR interactions. Longer exposure may highlight inefficiencies of the interaction paradigm on the desktop platform. For example, time on target is much slower than using a mouse to select a button. Although overall SUS scores were not statistically different, the Desktop Display group did observe an "excellent" overall score compared to Mobile VR and Video achieving "good". This difference may be explained by the Desktop Display group having more experience with PC interaction in general, whilst the video group could only watch and not interact with RPA. Overall, the usability scores were positive for RPA. This tool could be helpful as a revision tool for DBT lessons and a primer before movement is enacted. It could also aid other teaching frameworks that require pre-session study, like the flipped classroom.

2.8 Conclusion

This study explored the knowledge gained from using the Recovery Position Application as well as the usability of the system. The application was tested across three interaction and display configurations: Mobile VR, Desktop Display, and pre-recorded Video. Participants found the application easy to use, were optimistic about the use within an educational context, and knowledge scores were not mediated by display type or interaction paradigm. By analysing the interaction design requirements and hardware limitations, we have created an application that shows good usability at an early development stage. By translating the principles of effective observational learning into application features, the RPA demonstrates the effectiveness of mixed-reality devices (both immersive and non-immersive) in delivering virtual demonstrations of action sequences for observational learning. Such educational technology aims not to replace an instructor in DBT, but to supplement or inform users when an instructor is absent. This study makes no claims that virtual content is more effective than other media (for example, recorded videos of demonstrations). Instead, this body of work aims to show the variety of ways mixed reality tools can aid education and training for DBT. By expanding the strategies to deliver observational content, we are providing effective learning environments for a broader audience.

Through the development process of the RPA, we can recommend these guidelines for delivering demonstration content in 3D virtual environments:

- Optimise interaction controls factoring the limitations of the target platform to achieve effective usability. The same application may need significant development to be effective on a different platform (desktop PC, tablet etc.)
- Autonomy of navigation and information flow is important to users and so may impact the perceived usability over long periods of usage.
- However, for short demonstrations, the autonomy of information and navigation may not aid memory recall, if the demonstration is paced appropriately and showcases the action sequence clearly.
- Information should be broken down into small steps to aid recall. The amount of information provided per step will depend on the target user’s prior knowledge and

therefore, this should be accounted for in the design process.

- Many mixed-reality display types could facilitate virtual demonstrations. If the user has a variety of viewpoints to observe the required details, then a symbolic mental representation can be formed on both immersive and non-immersive displays.
- Combining visual demonstrations with audible commentary is an effective tool for information delivery. Establish context through the environment details and generate demonstrator animations for all spoken details.
- In-application viewpoints should provide both egocentric and allocentric perspectives to aid the user's mental rotation of enacted movements and to support the use of varied vantage points to gather action details.
- Graphical treatment should represent shape and form but does not need to be presented photo-realistically to be effective. For demonstrations, prioritise animation fidelity over the shape and form of the performers and environment.

2.9 Future Directions

The participant feedback has highlighted additional features and refinements that would improve the usability of the RPA navigation and information delivery:

- A repeat step function
- Breaking down of information into more steps
- Development of visual assets for the context of the scene
- Adjustments to viewing angles so translational movement is not required
- Establishing visual actions or elements that represent all audible information

The next stage in the development of RPA will be to design and implement these features for re-testing. Important to this will be to expand the participant pool in an ecologically valid setting to see the extent to which users can physically perform observed movements. A wider distribution and field testing with an instructor will be used to substantiate these initial findings and inform the future development of the application. Many more mixed reality devices may suit this type of demonstration. For example, augmented reality (AR) devices where the virtual demonstration can be superimposed on a real-world setting.

2.10 Limitations

Interviews used in this study had methodological weaknesses that should be addressed to avoid bias in results and gain more pertinent information. Although the semi-structured interview questions were developed with thought and focused towards specific goals, they require more stakeholder engagement in early development to refine. For example, discussing with developers any crucial areas to focus upon and performing a pilot to gauge the type of data these interview questions will produce. Additionally, the transcription and thematic coding require a formal structure to ensure that the analysis is data-driven and can be explored through various codes. For example, thematic template analysis ((Hamm, Money, Atwal, & Ghinea, 2019). The current study requires the researcher to draw out points based on goals which may lead to more bias and rigour in the holistic connections between interviews. These improvements to the interview methodology will aid the "trustworthiness of the study" (Kallio, Pietilä, Johnson, & Kangasniemi, 2016).

Participant perception of recovery position knowledge was used to measure previous recovery position knowledge. This approach needs to be more objective to measure actual knowledge of the recovery position sequence, as evidenced by the limited connection between previous knowledge and final recall. Therefore, an additional survey or task to measure the recovery position content and goals would better aid the analysis of what knowledge is gained through RPA.

Chapter 3

Experiment 01

Walking through virtual doors: A study on the effects of virtual location changes on memory

3.1 Chapter Summary

The motivation for work in this thesis is how to improve immersive virtual environments for training applications like the Recovery Position application. The key to many training scenarios is the transfer of information to the learner, so they can meaningfully interact with said information to develop their own schema for subject topics. As explained in the background section, virtual realities can traverse multiple locations whilst staying within one physical environment. Therefore approaches to information recall based on location are a potential paradigm for virtual reality training applications. This experiment investigates if segmenting word lists between virtual reality rooms can improve memory recall of information as observed in real-world studies (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016; Smith, 1982).

3.2 Introduction

A key design consideration for Virtual Reality (VR) experiences is how to pass on information to the user. In some cases, this will be an essential requirement of a VR tool. To illustrate, in education and training scenarios, users may need to retain a certain amount of information before they apply it to a given scenario. For example, recalling the action sequences and safety information before applying first aid. The incidental processing of a physical environment has been linked to key processes of memory and learning (Choi, Van Merriënboer, & Paas, 2014; Brooks, 2011). How we separate information between locations (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016), and how we travel between these areas can significantly impact what is recalled (Smith & Vela, 2001). This suggests that memory recall can be improved without the application of memory strategies by the user. VR is an ideal platform to apply and test these concepts, as users can traverse infinite locations without needing to leave their real-world rooms. This experiment explores whether separating information between VR rooms can aid recall, as observed in real-world trials (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016).

3.3 Background

Event cognition research emphasises the role of location in recalling a day's activities. Information is often categorised into events to sort through the daily stream of experiences. An event can be simply described as an amount of time within a location, that is perceived to have a beginning and an end (Zacks & Tversky, 2001). The mental representation and simulation of these events are called event models. An important construct of an event model is the spatial location and time of day that an event takes place (Radvansky & Zacks, 2011). To recall the events and associated experiences, one may first recall the spatial location of an event model. Event models act as a structure to recall the chronology and content of information that an individual is exposed to. Although the boundary of where one event model ends and the other starts can be established by any significant focus of attention within a location, notable changes in the environment are thought to define the placement of event boundaries, for example, walking through doorways.

Walking through doorways has been shown to aid and interfere with recall (Radvansky & Zacks, 2017). When information is shared between rooms (and therefore event models), both event models are activated on recollection, compete and cause retrieval interference (Pettijohn & Radvansky, 2016). This has been observed in experiments where participants are asked to recall or recognise objects they are carrying or have just carried between rooms compared to within a single room (Pettijohn & Radvansky, 2018). Conversely, walking through doorways can aid recall if information remains unique between rooms. This has been observed that presenting word lists between two rooms compared to one (but same walking distance) (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016). Much like breaking down one large list into two smaller ones, the event models act as memory hooks to smaller sets of information which can be brought together to make a larger body of knowledge. Similarly to event cognition work, environmental context-dependent memory research has also found evidence of improvements to recall of word lists when encoding occurs between multiple room locations (multiple contexts) (Smith, Glenberg, & Bjork, 1978; Smith, 1982). Showing more location changes (therefore separating a body of information between more locations) can aid memory recall and recognition displayed in various media. For example, video content (Gold, Zacks, & Flores, 2017), imagined narratives through reading, and recalling a word list from a desktop screen (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016). The recall enhancement appears to hold over longer durations, with improved information recall at one week (Flores, Bailey, Eisenberg, & Zacks, 2017) and two weeks (Radvansky, Andrea, & Fisher, 2017). Event frameworks will be robustly remembered after a month, but information within will have significantly reduced (Flores, Bailey, Eisenberg, & Zacks, 2017). This body of work suggests that recall can be improved by segmenting information between locations, whether within the physical world, imagined narratives or on media platforms.

Work that has observed an enhancement to recall by segmenting word lists between rooms has used real-world methodologies. Smith (1982), had participants learn four word lists in either a single room, between two rooms or four rooms. Participants would wait in a hallway between location changes. Recall significantly increased as room numbers increased. More recently, Pettijohn et al. (2016), exposed participants to two word lists in either a single-room or two-room condition. In each case, participants would walk toward the designated room location and have the first word list read to them. Participants would subsequently walk towards the following location for the second word list. Using multiple rooms (Participants walking through a doorway between word lists) significantly increased recall.

Virtual reality is a promising platform to test & apply theories of event cognition to information recall. Firstly, it affords a spatial understanding of a virtual location and a psychological sense of "being there" and thus creates the illusion that the user is in a new reality (Slater, 2014). Head movement is tracked in 3D space enabling a more natural scanning process of spatial cues compared to desktop displays. Allocentric (objects in relation to each other) spaces may also benefit from more immersive displays. When cave experts were asked a series of questions based on a 3D virtual map of a cave, tasks that involved finding map details and comparative measurements were answered more quickly and with greater accuracy using higher immersive displays (larger field of view, stereoscopic rendering turned on) (Schuchardt & Bowman, 2007). Therefore, an immersive display can aid the egocentric mapping of an environment, which will also help build the allocentric representation and, in turn, a cognitive map of the environment (Epstein, Patai, Julian, & Spiers, 2017). For example, users had a better spatial perception of a virtual architectural model when using an HMD (Paes, Arantes, & Irizarry, 2017) and increased recall of spatially placed information (Krokos, Plaisant, & Varshney, 2019). Secondly, content delivered on VR is not restricted by time and space, and therefore, can potentially use the principles of location-based event cognition to design experiences that facilitate recall. For example, segmenting a body of information between X number of rooms.

To avoid overlap of locational cues, studies that have used multiple rooms to encode lists of words have purposefully used perceptually different rooms. Lighting, general colour and function allow a clear distinction between environments (Smith, Glenberg, & Bjork, 1978; Smith, 1982; Glenberg, 1979). By keeping cues distinct, the variable of context is better established. A key attribute of a room is the prevailing colour. Colour may influence the memory of locations by enhancing attention and distinction of form (Dzulkifli & Mustafar, 2013). The more attention drawn to environmental stimuli, then the higher chance that the environmental piece is drawn into memory processes. This distinction of form can come from contrast (difference in light and dark attributes of colours) as well as the difference in the hue (red, green, blue etc.). Colour, compared to greyscale, can aid visual memory, which suggests that no colour would be detrimental to memory processes unless the contrast between objects were distinct. However, a high-contrast world devoid of colour might evoke other emotions due to being very unrealistic. If the distinction between hues and contrast may aid the memory of objects within structures, then using varied colours and lighting between locations may help establish well-defined contexts. Colour can also impact the memory of words differently between genders. For example, when recalling words in different coloured environments within VR, females showed better memory performance for high saturation colours and blue rooms (Nolé, Higuera-Trujillo, & Llinares, 2021). Males were not impacted by saturation but had significantly better memory performance in purple rooms. Both genders experienced the best recall when the lighting temperature reached 6500K. It is therefore, plausible that the colour of a room and its lighting can impact recall differently for either gender. These differences can also be influenced by cultural background. For example, Europeans prefer blue while the Chinese give more weight to the colour red (Marschall & Werner, 2018). More broadly, colour can mediate the emotional arousal of individuals. For example, when checking texts, those in a relatively negative mood in a red room worked through more text but with more errors (Küller, Mikellides, & Janssens, 2009). Increased speed and less accuracy, is a sign of emotional arousal. Conversely, colour can reduce emotional arousal. When surveying participants who read emotionally upsetting events, those that read these stories on pink paper, were less emotionally upset. Emotional arousal can be connected to words through colour and impact memory. When colour is congruent with an emotion associated with that word, higher recall is observed (Kuhbandner & Pekrun, 2013). For example, when the word danger is coloured red, positive words are coloured green.

These studies highlight a connection between participant attributes, colour, and emotion that can impact memory when there is congruence between these factors. This interplay will create tension for the design of virtual environments between location distinctiveness and impact on memory. Studies that have successfully observed enhancement to memory word lists when encoding through multiple rooms did so by choosing distinct locations (Smith, Glenberg, & Bjork, 1978; Smith, 1982; Glenberg, 1979) that varied lighting and aesthetics. Creating a contrast between rooms whilst limiting emotive words associated with room aesthetics should restrict this potential confound.

The incidental processing of a physical environment has been linked to key memory processes, including the episodic structuring of events over time and information retrieval. In such examples, the environment acts as a mental structure associated with information experienced within. By separating information between multiple locations or local areas, more information may be recalled compared to using a single location. If the separation of information between environments aids recall in the real world, then this concept could be a guiding blueprint to the design of better information recall within VR environments. To our knowledge, no articles can be found that explore the use of virtual reality in improving memory recall by physically walking through virtual doorways into room spaces.

H1: Within a VR environment, separating word lists between two rooms divided by a doorway will aid the recall of word lists compared to a single room.

3.4 Methodology

This study received approval from the Research Ethics & Integrity Committee, University of Plymouth. 29 (13 female) participants were recruited. The participants came from two distinct backgrounds, computer science and psychology undergraduate students. Participants were rewarded with credits towards class systems or free food. Participants took part in this experiment one at a time.

This experiment employs and builds on an existing methodology (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016) from event cognition work and was adapted to the VR platform. The same within-subjects design was used whereby participants would listen to two words lists in one of two conditions. Either a single room condition (1R), with both word lists in the same room or in a two-room condition (2R) with one word list in each room, separated by a doorway (Figure.3.2). After listening to both word lists in a condition, participants would take off their VR headset and be presented with a distraction task (a set of maths questions for 1 minute) and then asked to recall and write on a piece of paper as many words from the previous condition as they could. Once complete, participants would repeat this process for the alternate condition. Starting position (and therefore word list) and condition were counterbalanced across the trials. The recall of these two conditions was then compared.

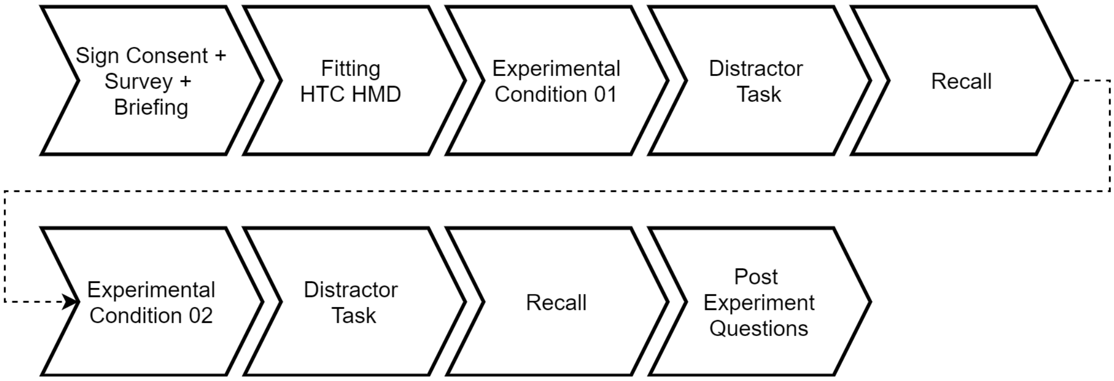
Sign Consent + Survey + Briefing
Fitting HTC HMD
Experimental Condition 01
Distractor Task
Recall
Experimental Condition 02
Distractor Task
Recall
Post Experiment Questions

Figure 3.1: Flow diagram showing the series of steps for this experiment's methodology.

For our experiment, the procedure was as follows: After signing a consent form, participants completed a demographic survey and were given a briefing on the upcoming task to explain that they would be listening to speakers and shown images of what the speakers looked like. These steps took at most 5 minutes. Participants were then fitted with the HTC Vive HMD (Head Mounted Display) (Vive, 2022) to provide the visual feed

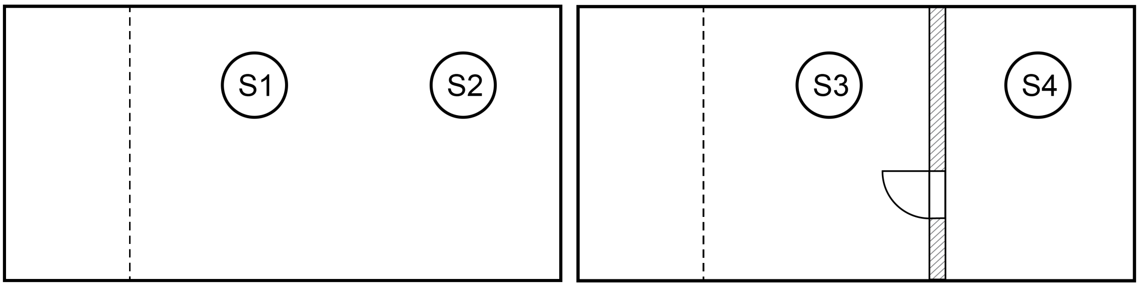

Figure 3.2: Room layout of the one-room (1R) (left) conditions snd two-room (2R) (right). S1-S4 = Speaker 01 - 04. Dotted lines indicated areas of visual interest.

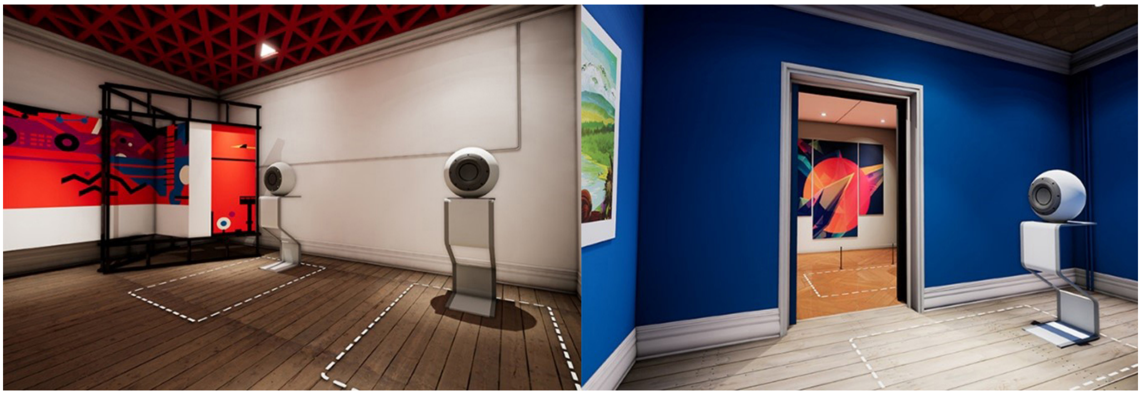

Figure 3.3: (Left) Image of the two-room (2R) condition. Speaker is on the right of the image and the doorway in the centre with a view of the second room. (Right) Image of one-room (1R) condition. Both speakers are placed within one room

of the virtual environment. The audio was delivered through headphones connected to the HMD. Participants started in the SteamVR Home scene (SteamVR, 2017) to check the proper fitting and visual clarity of the HMD. After the initial hardware check, the first experimental condition was loaded. Participants were asked to take a few moments (10 seconds) upon entering the virtual world to look around and orientate themselves. After orientation, the participants were instructed by the experimenter to approach the speakers and remember the delivered words best that they could. Orientation and listening of both word lists took 2-3 minutes per condition. After listening to both speakers, the participants were asked to remove the Vive HMD. Participants were then given a set of maths questions on a sheet of paper to work through for one minute to encourage forgetting and reduce the use of short-term memory strategies like the repetition of the word list. The experimenter then asked participants to recall as many words as possible from the virtual environment by writing them down on the back of the maths questions paper. This procedure was then repeated for the following condition.

After both conditions were completed, the participants were invited to take part in a survey that asked if any memory strategies were used and if they perceived themselves walking between rooms in the two room condition. These survey questions were used to identify approaches to memorising words and help establish if participants were cognitively processing multiple rooms in the two-room condition. A memory strategy was defined as a technique that would aid long-term recall. For example, creating a story from the presented words during encoding. Strategies that aided short-term recall were categorised as not using a memory strategy, as the distractor task between encoding and recall should minimise any benefit to recall from such strategies.

Four sets of 10 words were required for this experiment, 40 words in total. In line with previous experiments (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016), the MRC Psycholinguistic Database was used Database (1997) to generate the words. Each word was one syllable, five letters long, and ranged in word frequency from 20 to 103 per million. 226 words were generated, and then 40 of these were randomly selected for this experiment. These 40 words were randomly divided into four word lists. A subjective check was made to ensure that each list's words were not too similar or held a strong implicit association (Word list used can be found here (Watson, 2023)). A word list was then assigned to each speaker. Each speaker would deliver their assigned word list in a random order for each participant. Two virtual environments were developed within the Epic Unreal Engine 4 (Engine, 2004) (Figure.3.3) for the conditions of this experiment: A single room and a two-room environment. The overall virtual area used for each condition was the same, and the distance between the speakers was 2m for both conditions (Figure.3.2). However, the aesthetics used within each room were different. For the 2R condition, a dividing wall and doorway connected the rooms that opened automatically once the first word list had been listened to. The aesthetics of the virtual environments were different for each room. The goal was to have distinct memory cues for each room to aid attention on the environment and emphasise memory hooks for each. The concept of an art gallery was used to distinguish between rooms. In such buildings, shape, form, and colour are used to create engaging, aesthetically pleasing rooms. Therefore, in such rooms, it would not seem strange to have individual room colours and images adorning the area. Each environment has an area of "visual interest" (See Figure.3.2) to act as a landmark for room representation. This would be a large image on a canvas. Each room used a different colour. In the 1R condition, the walls were white, but the ceiling was distinctly red. In the 2R condition, the walls of one room were a dark grey and the walls of the adjacent room bright blue. This was to provide a distinction between rooms in the 2R condition.

In this experiment, word association with environmental colour is thought to be limited. Firstly words were randomly selected into lists and reviewed to appear commonly within the English language. Secondly, as words were spoken to the participants, no colour was directly placed upon a word. The bright blue colour may give females an advantage in this room for word memorisation. However, this experiment has a good representation of both males and females which should allow any colour differences by gender to be revealed in the analysis.

In both conditions, there were two speakers that when approached by the participant, voiced ten randomly selected words with a delay of one second between each. Participants could physically walk within the virtual world between word lists and through rooms to reflect the same navigation method used in previous work. To adapt this methodology to the VR platform, a human male was recorded speaking the word lists. These voice files were delivered to participants through virtual speakers. Speakers would automatically deliver word lists (A one-second gap was used between each word) when participants approached them. A difference in the methodology used in this experiment was that the distraction task and the free recall occurred in the real world and not the virtual, acting as a third room for recall. This approach was chosen as 1) previous work ((Smith, 1982) - experiment 3) suggests that recalling a word list in the same room as the encoding will elicit a context reinstatement effect, whereby the environmental cues exposed during encoding aid recall. However, context reinstatement is significantly reduced when using multiple rooms. Therefore recalling in the same room for both conditions would give a context reinstatement advantage to the single room condition and is not the subject of this study. Using a different context (room) for recall for both conditions will control for any context reinstatement effects. 2) it allowed a more direct replication of previous distraction and recall tasks.

For a video walkthrough of the virtual conditions, refer to "EXP 1" work presented in the supplementary material (Watson, 2023)

3.5 Results

Of the 29 participants recruited for this experiment: 15 participants presented increased recall in the 2R condition, 12 participants presented an increase in recall in the 1R condition, and 2 participants saw no change in recall between conditions. Mean recall for the 2R

Table 3.1: Participant Demographics

Demographic Information				
Age	18-24	25-34	35-44	45-54
	21	4	3	1
Gender	Male	Female		
	16	13		
Used VR before?	Yes	No		
	20	9		

Table 3.2: Post experiment survey

Survey Questions	Yes	No
Did the environment with the doorway feel like two separate rooms?	27	2
Did you use any memory strategies to help remember the words?	14	15

condition 4.59/20 (SD = 1.88). The mean recall for the 1R condition was 4.45/20 (SD = 1.96) (Figure.3.4). Recall data for speakers 2 and 3 (Figure.3.2 for speaker layout) were not normally distributed as assessed by Shapiro-Wilk's test (p = < 0.05). Therefore, a Wilcoxen signed-rank test was run to analyse the degree of recall difference between the 2R and 1R conditions. Statistically, there was no significant median difference in recall between conditions, Z = .503, p = .615, Hedges g = 0.08 90% CI [-0.37 – 0.53].

In post-experiment questionnaires (Table.3.2), participants were asked if they used a memory strategy when encoding the word lists. On analysis, no significant effect of memory strategy use was found. To test for recall differences between individual speakers, a Friedman test was run. Word list recall was similar across speakers 1-3 (Mdn = 2.00), with an increase at speaker 4 (Mdn = 3.00). These differences were not statistically significant, 2(3) = 6.536, p = .088.

To test for recall differences between genders, a Kruskal-Wallis test was conducted. The results indicate a significant difference ($\chi^2(1) = 7.248, p = .007$), with females recalling more words ($M = 10.92, SD = 2.84$) than males ($M = 7.50, SD = 3.01$). This difference can be seen in Figure.3.5 where there is a clear increase in recall for females across speakers one to three and a trend (slight overlap of error bars) for better recall on speaker four.

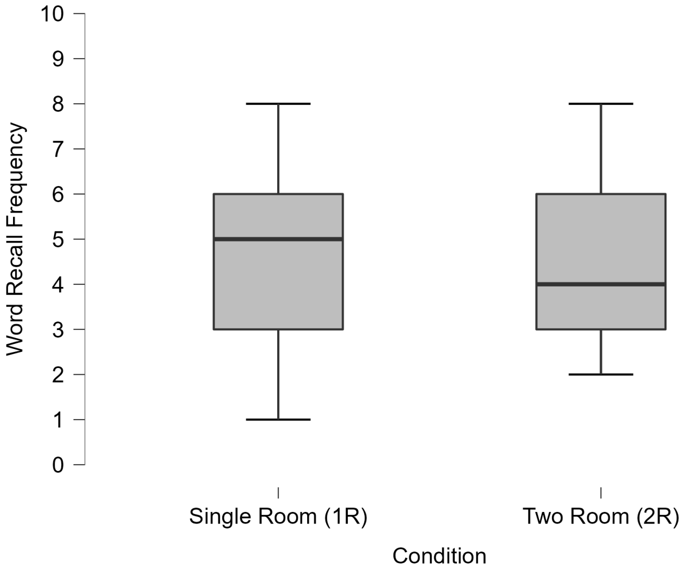

Figure 3.4: Boxplots of word recall across single-room (1R) and two-room (2R) conditions. 2R has a median value of 4 words recalled compared to a median of 5 for the 1R condition. However, the interquartile range and variance are similar, suggesting no significant difference between conditions. This is confirmed through significance testing

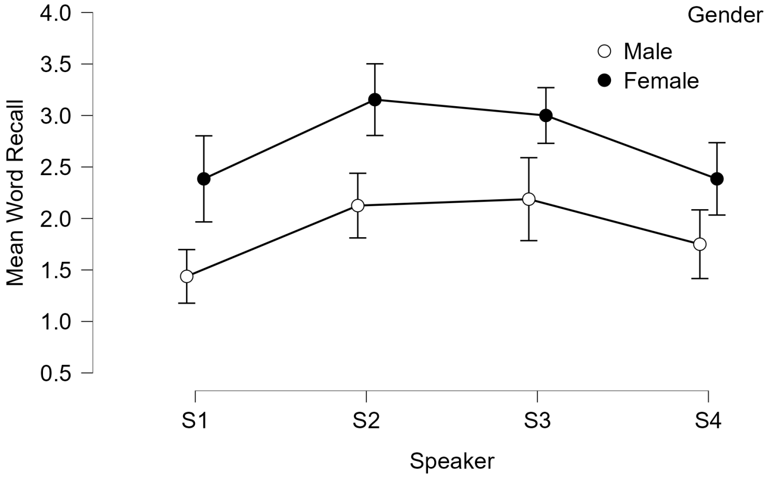

Figure 3.5: Graph to show the difference between recall at each speaker for each gender. The single-room condition contained S1 and S2, and the two-room condition contained S3 and S4. Error bars suggest that the difference in recall between speakers for each gender is not significantly different, while females significantly outperformed males at the free recall task.

3.6 Discussion

Following on from real-world studies that suggest segmenting information between multiple rooms can aid recall (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016; Smith, 1982), we investigated if the same effect could be observed within VR environments. Even though participants perceived the virtual two-room condition as two different rooms, no significant difference in recall was observed when segmenting information between multiple VR rooms.

Improving recall of word lists by segmenting the encoding between different locations has been replicated in real-world methodologies. There are instances when replication did not reach significance (Smith, 1984). However, there was still some difference in recall between using a single room and multiple rooms for encoding. Meta-analysis of environmental context work suggests that segmenting information between contexts (rooms) to learn word lists is reliable with a moderate effect size (d = 0.45). This, combined with event cognition work that has observed enhanced recall from segmenting word lists between event models (through narrative, locations and other media), suggests that it is a replicable effect across media platforms for both within and between subject experimental designs.

The overall recall was significantly higher for females as opposed to males. This recall is significantly higher for speakers one to three, with a trend of higher recall for speaker four. A defining attribute for females compared to males was that the females all studied a psychology course and the males on a games development course. Therefore, psychology students may be more practised with words and memory cognition, which aided their recall. More likely, this is not a gender difference but an academic background distinction aiding those from a psychology domain.

The room's colour choice was guided by the goal of making each room distinct. The aim was that the context of an art gallery would allow participants to accept distinct colours as part of the aesthetic and aid environmental cues for each room. A potential confound that aids females would be room colour. Speaker four would have been in the room coloured with blue walls. According to findings from Nole Nolé et al. (2021), this should have given a memory advantage to females. However, the word list delivered by this speaker was the only one not to observe a statistically different recall result between females and males. However, this previous work had 50 females participate as opposed to the 13 in this study.

This may mean a larger sample size would show the effects of colour. The recall between speakers for females and males was not statistically different. This suggests there was no advantage to recall for any specific speaker. Since each speaker has an assigned word list, this suggests that there was limited interaction between a word list and a room design that either aided or suppressed recall. However, this interaction cannot be investigated further, which leaves the possibility for a confound of a small effect size hidden by the interaction of room design and word list.

This experiment closely followed past methodologies from event cognition work (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016), applied to a VR platform. Interestingly, in comparing this experiment to previous results, we observed a non-significant difference between conditions, and average recall across this study was lower. The two-room condition observed a 9% (2 words) drop in the comparative recall, and the single-room condition 5% (1 word) drop in the comparative recall. Additionally, the effect of memory strategies was also not significant. The use of internal strategies should supersede the effects of location on memory and show greater recall independent of how the information is segmented between rooms (Smith, 1985). This suggests more interference in the encoding or recall process and potentially greater difficulty in applying memory strategies. Our current experiment may have limited statistical power due to the sample size (N = 29). Previous work Pettijohn et al. (2016) observed a moderate effect size d = 0.629 for improving recall by separating word lists between locations. Analysis of power based on this work (Westfall, 2015) estimates that approximately 55 participants are required to obtain the statistical power of 0.80. Therefore, this current experiment may not be sensitive enough to observe the same effect. In the presented experiment, we observe a smaller effect size (d=0.07). Given that overall recall is lower than in previous work, memory strategies did not appear to increase recall significantly, and a smaller effect size was observed. These results suggest that an element of the methodology or the VR platform interferes with the use of event models to structure and aid the recall of word lists.

One methodological explanation for no effect of separating information between rooms is that previous work Pettijohn et al. (2016) had participants recall words in the same room as they listened to the word lists. In this current experiment, the participants removed their HMD to perform the distraction and recall tasks. This approach was used to make sure that there was a different environment for recall so no context reinstatement effects would give an advantage to the single room condition. This approach should not interfere with using multiple locations to encode word lists. However, some recent work suggests that there may be a more fundamental reduction in recall when recalling words in the real world that were memorised in VR (Lamers & Lanen, 2021). Participants performed memorisation tasks at a virtual desk or real-world desk. They were then called back 24 hours later to recall and recognise this information in the same or opposite reality. Those who learnt information in the VR context recalled significantly less in the real world than in all other conditions. However, more work is required to establish the extent of this effect in larger groups and different interval times before final recall. It suggests that simply replicating an environment may not engender the same cognitive response as the real world. Based on this work, developing a unique recall environment within VR would minimise the effects of context reinstatement while also avoiding any mediating influence of removing the headset on the final recall.

The navigation and interaction method for this current experiment were chosen to be as natural as possible to reflect real-world methodologies being replicated. Each room was distinct in colour, shape, and décor. The culmination of these attributes was to simulate the belief that participants were passing into separate rooms and minimise overlap in environmental cues that could cause competition at recall. Participants overwhelmingly felt as though they were passing between two separate rooms in the two-room condition. This suggests that the broader, global details were sufficient for event model creation. However, some local details overlap between conditions. The speaker design and voice used to deliver word lists were the same at each information point. From an event cognition perspective, event models can both help and hinder the recall of information (Radvansky, 2012). When experienced information is clearly segmented between separate event models, then there is potential for better recall. If event models share information, then interference occurs and recall is reduced. Environmental context work (Smith & Vela, 2001) suggests a key driver for larger effect sizes is variation in the delivery of information, for example, a different experimenter delivers the word lists. It is plausible then that using the same speakers and voice for the word lists was enough to generate interference from competing environmental cues within the conditions. Potentially local information variation is as important as global room differences for this study.

Relating more broadly to environmental context-dependent work is the phenomenon of context reinstatement, where returning to a context exposed during encoding can aid recall. For example, enhanced memory of information when recalling in the same room that the information was learnt. Although distinct from multiple context segmentation of information, they share fundamental drivers, linking encoded information to a location. Context reinstatement has had difficulties with replication in VR. Wälti et al. (2019) could not replicate enhanced recall from context reinstatement across simple background contexts (pc screen with colours and words), visual richness (using landscapes as contexts) and immersion levels (HMD vs none). Throughout these trials, participants were sitting down and had limited interaction with the locations. Another well-studied phenomenon when walking through doorways is reduced recall and recognition of information shared between the rooms. For example, knowledge of an item that has been picked up or placed down in a previous room. This is known as the location updating effect. The interference is caused by competition between the multiple event models assigned to each room. Although considered a reliable effect, this was not replicable within VR, using previously established methodologies (McFadyen, Nolan, Pinocy, Buteri, & Baumann, 2021). This experiment used passive navigation (movement through rooms was automated) and each room was purposefully identical. And yet Shin et al. (2021) was able to evidence of context reinstatement for free recall of collected objects within two VR contrasting worlds. A key difference in Shin's work was developing an understanding of the world through exploration, objectives, and distinct interactions. The rationale is that if an individual has the experience and a developed schema of a location, they will have a deeper mental representation to build cues for recall. After familiarisation with the locals, the participant's task was to search for objects and rate their usefulness before a surprise recall test. This approach would most likely use more tools to aid memory than just contest reinstatement. It may encourage a method of loci style memory strategy on the participants as they use local environmental constructs to associate object names. This has shown to be effective in VR (Krokos, Plaisant, & Varshney, 2019) even when the participant are not explicitly taught the method of loci technique (Huttner, Qian, & Robra-Bissantz, 2019). Additionally, it has been evidenced that a simple search for objects within a VR room without the intent to memorise objects is more effective for recognition tests and object location knowledge compared to being shown the objects in a room with explicit instructions to memorise them (Helbing, Draschkow, & Võ, 2020). Shin's work, therefore, shows that familiarisation and schema development of an environment can strengthen cues used for context reinstatement or aid the efficacy of loci-based memory strategies without the need for an individual to apply the strategy explicitly. However, regarding more broadly the use of environmental cues to help structure and recall information, it may suggest that if an individual is not able to create an internally understood context for each location then there is a chance that many locations can be merged into a higher level context on recall. For example, putting on a VR headset and entering a new reality overrides the individual contexts of each virtual location. Although the virtual environments used in our experiment are realistic in style, the colours, shapes, lighting, and composition are synthetic and in high contrast to the laboratory that the participants start in. This may be amplified by the difference in the field of view that the HTC Vive HMD provides. The HTC Vive has a field of view of 110 degrees. Although wide, it is less than natural vision (approximately 180 degrees), with the borders of the lenses viewable in the periphery of the user. This will give the effect of wearing big goggles. If participants are not given sufficient time to familiarise themselves with the VR environment and the technology, then environmental cues of individual rooms may be superseded but the general context of the VR platform. Therefore, a longer familiarisation time within VR could help the establishment of environmental cues and would control for technology experience variation within the sample (Lopez, Deliens, & Cleeremans, 2016).

3.7 Limitations and Future Work

Although the design of separating information between two virtual room locations did not enhance memory recall, this has identified future work that would help to clarify if the effect was superseded or too weak to be pragmatic as a design approach to information delivery on VR platforms.

It is unclear if the act of removing the HMD interferes with real-world recall. Repeating the same work with free recall tasks inside of the VR world would help address this interesting question. Future work could also consider replicating real-world segmentation of information between rooms to compare directly against the VR platform.

A familiarisation period should be established to make sure that participants are sufficiently acclimatised to the VR world and control for potential population variation of experience with VR platforms.

Larger sample size is suggested to increase the sensitivity of this methodology to help identify confounding effects or the potential that the effect of multiple rooms aiding recall of word lists is weaker within VR platforms.

The strength of the VR platform is that the virtual space can be adapted or kept consistent between methodologies. The difference between locations, and in particular local foci of attention (speakers aesthetics) could be exaggerated for more distinct cues to associated delivered information.

To avoid word lists being a confounding variable, they should also be counterbalanced so that a random word list from a selection is presented to a participant at each speaker. This will remove any potential confound that a word list mediates recall, and so allow deeper analysis of other factors against recall.

3.8 Conclusions

We investigated if separating word lists between two virtual rooms could help recall compared to a single virtual location. Results indicate that there is no significant benefit to recall when separating spoken information between virtual environments. However, this experiment also found no significant impact of memory strategy when used by participants, and recall rates were lower than in previous work performed in the real world. We have discussed potential improvements to the methodology used that may interfere with the enhancement of recall when segmenting word lists between rooms as observed in real-world studies. VR is a great tool for information delivery, but applications that value retention of spoken information may need to account for a possible reduced recall when using VR that could mediate information assimilation compared to real-world counterparts.

Chapter 4

Experiment 02

Walking through Virtual doors: Comparing virtual and real-world location changes on memory

4.1 Chapter Summary

This chapter describes the methodology, results and analysis of experiment two. Experiment two builds on experiment one as it shares the same premise but addresses some identified limitations. Participants must listen to two 10-item word lists held within one room or segmented between two rooms. Participants are required to recall as many words as possible from these lists. In this experiment, a repeated measure design is used where participants listen to these word lists within a single-room and two-room conditions in both virtual reality and the real world. This is to both validate Pettijohn et al. (2016) approach through closer replication and compare against virtual reality to see if the technology or implementation may alter participant recall performance.

Some limitations of the virtual reality implementation are also addressed. Firstly, an acclimatisation period is added to teach participants the controls and familiarise them with virtual reality. Secondly, the recall is now verbal and occurs in the room of the last word list. This is the same approach used by Pettijohn et al. (2016). In this study, participants interacted with both doorway and speakers to navigate between rooms and play a word list.

4.2 Method

4.2.1 Participants

30 participants (17 female) were used for this study. A mixture of computer science and undergraduate psychology students from Plymouth University participated. Participants were rewarded with credits towards class systems or free food.

4.2.2 Grouping

This experiment used a repeated measures design. Participants were tasked to listen to two 10-item word lists for recall afterwards. This was repeated in four different conditions. In one condition, the word lists were placed in the same room (See Figure.4.1) and in the following condition, spaced between two separate rooms, with a doorway in between. This created a Single Room and Two Room levels. These conditions were repeated in the real world and virtual world. Therefore, there were four conditions in total:

- Virtual World Single Room (VW1)
- Virtual World Two Room (VW2)
- Real World Single Room (RW1)
- Real World Two Room (RW2)

4.2.3 Interaction and Delivery

Words lists were delivered through speakers that the participants operated. To start a word list, the participant tapped a mobile phone screen in front of a speaker (See Figure.4.2). This interaction method was the same in both virtual and real-world conditions. The speaker would then deliver 10 words spaced 1 second apart. The MRC Psycholinguistic Database was used (Database, 1997) to generate the words. Each word was one syllable, five letters long, and ranged in word frequency from 20 to 103 per million. Eight sets of ten words were selected for this study (one for each speaker).

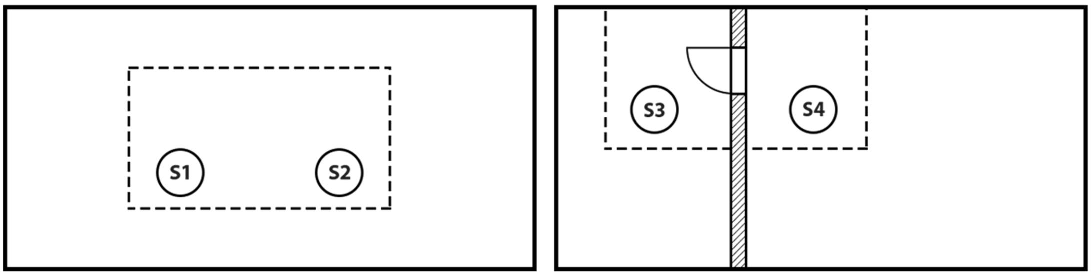

Figure 4.1: Showcasing room layout of the single-room (Left) and two-room (right) conditions. S1-S4 = Speakers. The dotted line represents the navigatable area that a participant can explore in the virtual world

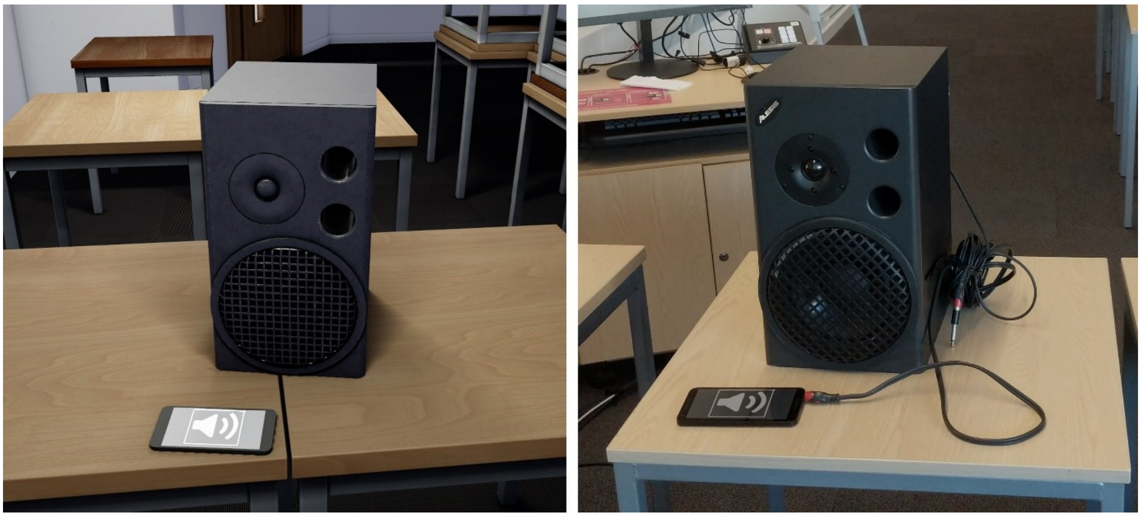

Figure 4.2: Above images show the speakers that deliver the word lists (left – Virtual World, right – Real World). Participants would tap the mobile phone screen in front of the speaker to activate the word lists.

4.2.4 Procedure

Participants were told how the speakers could be activated (tap on the mobile screen) to deliver a word list. For each condition, participants were asked to listen to each speaker twice and remember the information delivered as best they could. For two-room conditions, participants would have to open the door before moving through the doorway. For all conditions, pairs of speakers were positioned the same distance apart (205cm). After listening to both speakers, participants were given a distraction task for one minute. This would be a set of maths questions delivered by the investigator or counting backwards by N challenge. Participants then verbally recalled as many words as they could from the speakers within that condition. Post virtual conditions, participants were interviewed about their experiences.

The order of rooms, starting speaker, distraction task, and reality of the environment (real-world or virtual) were counterbalanced. There was a 48-hour gap between real-world and virtual conditions. All four rooms across the conditions required 40-60 minutes to complete.

After participants had completed their recall task in the virtual conditions, they were invited to take part in a semi-structured interview (5 – 10 mins). There were three areas of investigation for this interview to help assess if the virtual implementation sufficiently replicates real-world conditions. Firstly to investigate if the VW2 condition was perceived as two separate areas. If participants do not perceive the separet rooms, they are unlikely to cognitively generate an event model for each. Secondly, if participants felt as though the tutorial facilitated enough training to interact with and acclimatise to the virtual world. If participant's attention is focused on how to interact or the novelty of being in an immersive virtual world, then they will have fewer cognitive resources to take part in the memory activity. Lastly, the interviews were used to investigate if participants felt that the VR hardware or world was distracting from the memory task given. The interview was a series of closed questions ("In the environment where you had to walk through a door to access the next word list, did it feel like two separate rooms?") followed by open ones to encourage reflection on the justification behind the closed questions ("Can you explain why it felt like two separate rooms?"). The researcher could adapt the open questions to better explore these responses. The researcher would input responses into an online form whilst administering the interview. The closed questions received either yes or no answers and so could be analysed as to whether or not the participants agreed with the statements. The open questions were analysed as to reasons for either agreeing or disagreeing with the statement.

4.2.5 Learning the Virtual Controls

Participants engaged in a five-minute tutorial for the virtual conditions before any testing began (Figure.4.3). The tutorial scene aimed to teach the participant the experiment's controls by allowing them practice of interacting with doors and audio sources. One activity in the tutorial required participants to throw virtual balls at moving targets. However, participants had to press a button behind a door to gather ammunition for this task. In this way, participants practised opening doors and natural interaction with objects. The tutorial helped to acclimatise participants to the HTC Vive hardware and facilitated testing per participant for any tracking, visual, or audio errors pre-experiment.

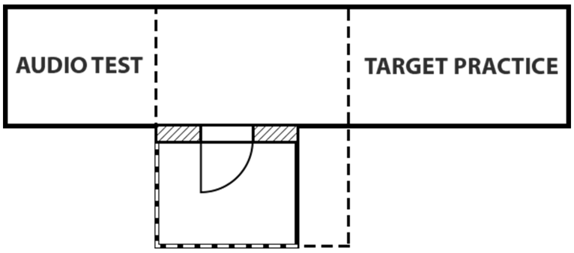

Figure 4.3: Above floorplan shows the layout of the tutorial room. The dotted line represents the navigatable area a participant can explore in the virtual world. Here participants can test interaction with audio sources, using doors and throwing projectiles at targets to practice virtual controls and interactions.

4.2.6 Hardware and Software

3 virtual rooms were created within Epic Unreal Engine 4.20 (Engine, 2004); a single-room, two-room, and tutorial. The aesthetic of these rooms was based on their real-world counterparts (See Figure.4.4). However, the lighting and floor plans were different. Similarities in aesthetics reduce visual variations between rooms that may influence memory recall. The difference in floorplan will help identify each room as an individual space. HTC Vive HMD (Vive, 2022) was used to display the virtual environment. The HTC Vive controller was used to interact with objects (doors, audio sources etc.). This allowed natural movement (walking) and interaction (using hands as points of interaction) within the virtual spaces. Participant verbal recall was recorded on a laptop microphone.

For a video walkthrough of all virtual levels, refer to "EXP 2" work presented in the supplementary material (Watson, 2023)

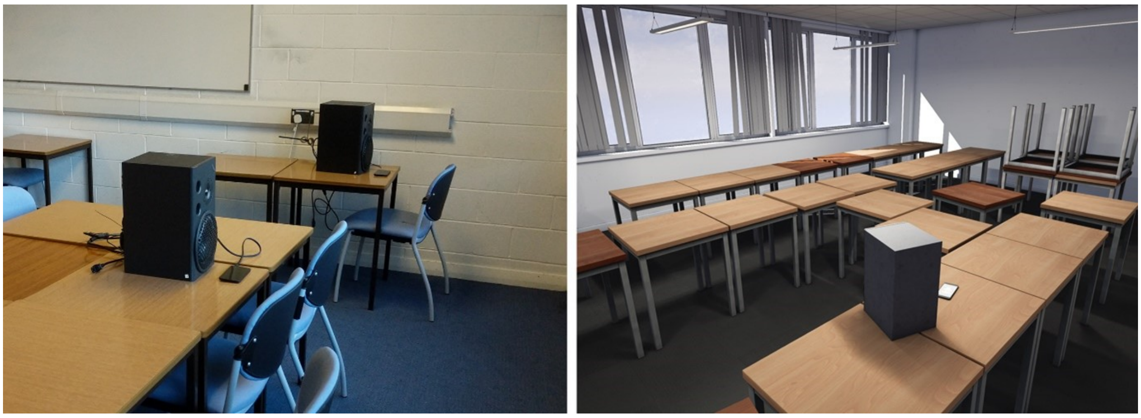

Figure 4.4: Images to show visual aesthetics of the real world (left) and virtual world (right) for the single room conditions.

Real-world word lists were delivered by tapping on the screen of a Nexus 5X smartphone when the word list application was loaded. An Alesis M1 Active Reference speaker was used to output the audio from the Nexus 5X. The aesthetics of the Nexus 5X and Alesis speaker were copied for the virtual conditions (See Figure.4.2). In the virtual conditions, participants used over-ear headphones connected to the HTC Vive audio jack to listen to sound content. Virtual word lists were played in stereo, and adjusted the volume on the left and right sides to match the relative position and rotation of the participant's head. This simulates listening to an audio source in 3D space.

4.3 Results

Post-experiment, participants (N = 30) were interviewed on their acclimatisation to the virtual world and the interactions required for the experiment. 29/30 of participants perceived there to be two distinct rooms in the VW2 condition. When prompted, there was interesting variation as to the extent of this perception. One participant described, "definitely, it felt like a different space.... I used the visualisation of the other room to help my memory...". Another participant mentioned, "Yes, but moving from one room to another did not feel as significant as normal. It felt like a door, but I knew I was not in another room."

29/30 participants felt that the tutorial enabled them to understand the required controls of the virtual environment. Participants' comments were positive, describing "Definitely" and "I understood how to use the controls and navigation.."

27/30 felt they were used to the virtual world by the end of the tutorial. One participant highlighted awareness of both realities simultaneously: "It's hard to get used to, realistic but feels artificial. Still very aware of the real world." Another highlighted less acclimatisation compared to the real world "More than I was, but not as used to the real world". Additionally, it was pointed out that not being able to see own body felt odd "Felt weird that I could not see my own body or check my watch", but perceived this not to interfere with the study "Even though the lack of body was weird, I did not feel it interfered with the task"

23/30 were not distracted by the virtual world or the hardware worn when performing the memorisation tasks. Two participants found the novelty of being within an immersive virtual world distracting. "I was curious about the environment. The novelty of technology

"How amazing is this, I was thinking", "A little bit distracted by wanting to explore the virtual world. ".

Two participants felt the hardware was distracting "It felt enclosed, you could notice the headset on… Overall a bit distracted" "Because it's VR, it is more to process your brain… the environment made it easier to remember, but the technology makes it harder." Two participants felt unsettled and distracted by the replication of a realistic setting inside of VR. "Classrooms looked eerie. The experience was a bit weird since it was very much like real life.", "You know you are not in real life, but you have never seen or interacted with this world before."

For recall analysis, one participant was deemed to be an outlier due to almost scoring perfect recall. Two participants did not attend one of their experiment sessions. The closest match approach was used to add data due to these dropouts. In repeated measures designs, the closest match approach uses "donor" data from another participant who scored similar results on a measure taken at a different time point. This has been analysed as an effective imputation with a sample size of 25-50 (Elliott & Hawthorne, 2005).

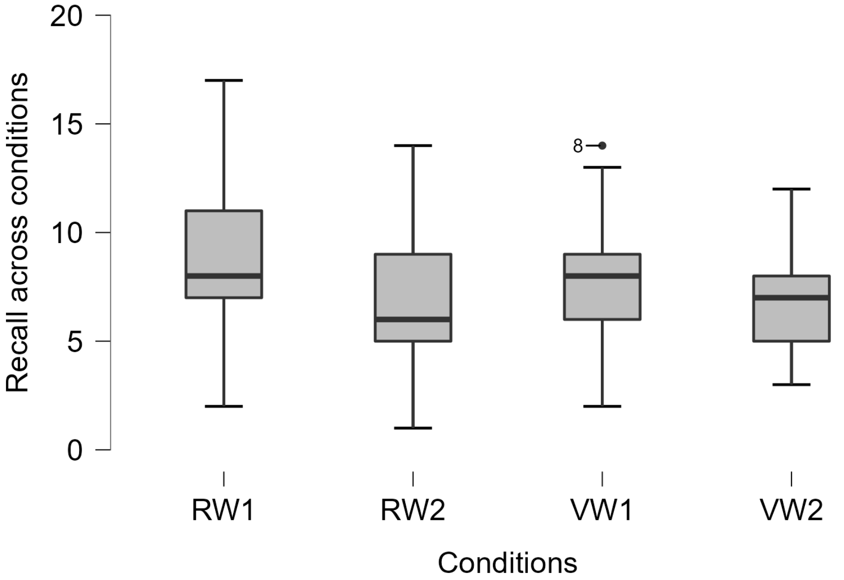

Figure 4.5: Boxplot to show the distribution of recall between conditions. RW1 had the highest average recall but also the highest variation.

Figure.4.5 shows that RW1 (real-world reality with one room) had the highest recall but also much more variation in the recall than other conditions. RW1 did observe significantly higher recall when paired against all other conditions (See Table.4.1). No other paired conditions presented significantly different results between the mean recall of word lists.

A two-way repeated ANOVA was conducted to explore the main effects on recall for

Paired Samples T-Test

Pair	t	df	p
RW1 - VW2	3.617	25	.001
RW1 - RW2	4.756	28	.000
RW1 - VW1	2.574	28	.016

Table 4.1: Shows that RW1 (Real World Single Room) facilitated significantly higher recall than all other conditions.

condition RW1. A statistically significant two-way interaction was found between the type of reality (real or virtual) and the number of rooms (one or two), $F(1,28) = 4.77, p = .037, \eta_p^2 = .15, 90\% CI[0.01, 0.51], \eta_G^2 = .01$.

When the number of rooms was analysed as a simple main effect, RW1 presented a mean increase of 1.931 compared to RW2. This was statistically significant, $t(28) = 4.76, p =< .001, 95\% CI[1.10, 2.76], Hedges'sG_{a}v = .55(CI90[0.33, 0.79])$. However, the mean difference in recall scores was not significant for the virtual conditions (VW1 vs VW2) with a difference of $0.62, t(28) = 1.242, p = .225, 95\% CI[-0.40, 1.64], Hedges'sG_{a}v = .23(CI90[-0.8, 0.54])$. This suggests that the Room Number factor was not a main effect. Similarly, when reality was examined as a simple main effect, there was a statistical mean increase of 1.48 for RW1 compared to VW1, $t(28) = 2.574, p = .016, 95\% CI[0.30, 2.66], Hedges'sG_{a}v = .55(CI90[-0.8, 0.54])$, but not RW2 compared to VW2, with RW2 achieving a higher mean recall of 0.172, $t(28) = 0.763, p =< .001, 95\% CI[-0.99, 1.33], Hedges'sG_{a}v = .06(CI90[-0.38, 0.26])$. This suggests that the type of reality was also not a main effect. Therefore, the significant increase in recall for RW1 cannot be explained through trends from main effects but an interaction between reality and room number.

The significant increase in RW1 required further analysis to investigate the cause. Therefore, post-experiment, participants were contacted by email to explore what memory strategies were used when encoding and recalling the word lists. 16/30 participants responded. This data was used to group participants into those that used a memory strategy and those that did not. For this categorisation, a memory strategy was defined as a technique that would aid long-term recall. For example, creating a story from the presented words during encoding. Strategies that aided short-term recall (for example, repetition of words) were categorised as not using a memory strategy, as the distractor task between encoding and recall should minimise any benefit to recall from such strategies. For those that used a memory strategy, two participants attempted to associate words with the environment, four participants created a story from the words, one participant associated words to own body actions, and two created associations between word pairs whilst also attempting to link words to images. These were counted as memory strategies as they all allowed participants to recall the story, category, image, or action to bring to memory the associated words.

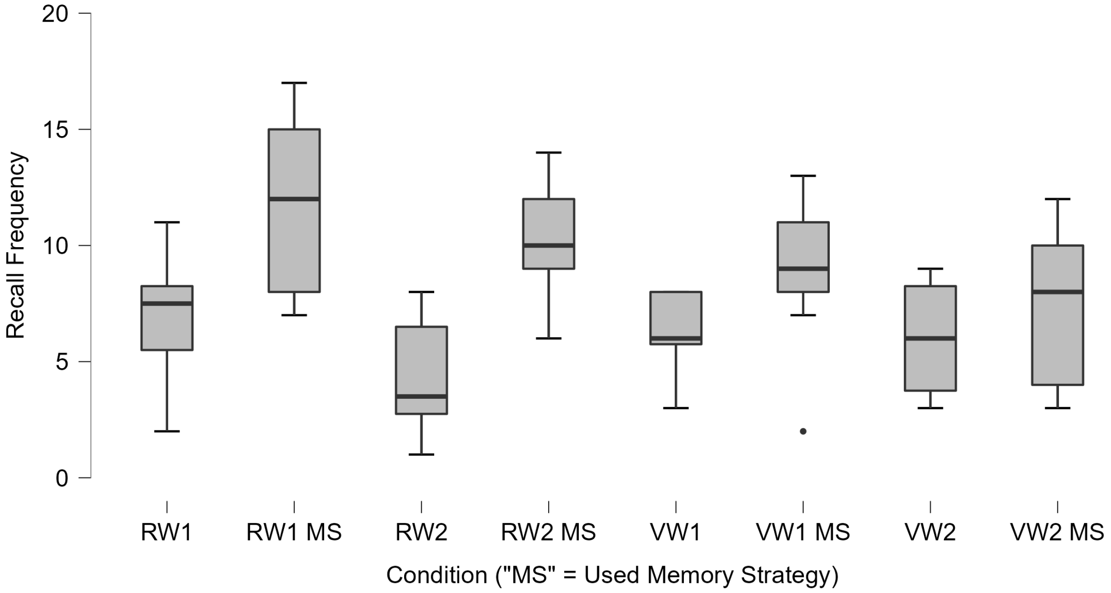

Figure 4.6: Box plot visually shows the improvement in memory recall when using a memory strategy (Condition + MS). In real-world conditions (RW1, RW2) the effect of a memory strategy is greater than the virtual conditions (VW1, VW2), with least improvement to recall when using a memory strategy for VW2 compared to all other conditions.

Subsequently, the data (N = 16) was analysed by comparing those that used a memory strategy and those that did not. A three-way mixed ANOVA was conducted to compare the effects of reality, the number of rooms, and memory strategy. The interaction of these variables was not significant, $F(1,15) = 1.921, p = .186, \eta^2_p = .11, 90\%CI[0.00, 0.36], \eta^2_G = .01$. Interaction between the type of reality and use of memory strategy was statistically significant, $F(1.15) = 7.430, p = .016, \eta^2_p = .33, 90\%CI[0.04, 0.55], \eta^2_G = .08$. Also, the use of memory strategy as a between-subject effect was statistically significant, $F(1.15) = 10.972, p = .016, \eta^2_p = .42, 90\%CI[0.10, 0.61], \eta^2_G = .31$. The large effect size of memory strategy indicates that it could be a key driver in aiding recall for real-world conditions. To verify this, independent T-tests were performed on each condition grouped by memory strategy. Results show a significant increase in mean recall when using a memory strategy in all conditions except VW2 (See Figure.4.6, Table.4.2). Real-world conditions saw a greater increase in mean recall compared to virtual conditions when using a memory strategy. Together, these results suggest that real-world conditions better facilitate the use of memory strategies.

Pair		Descriptive Statistics			Independent samples t-test		
Cond.	Mem Strat	Mean	N	SD	t	df	p
RW1	No	6.88	8	2.85	-2.91	15	0.011
	Yes	11.78	9	3.93			
RW2	No	4.38	8	2.67	-4.54	15	0.000
	Yes	10.22	9	2.64			
VW1	No	6.25	8	1.75	-2.13	15	0.050
	Yes	9.00	9	3.24			
VW2	No	6.00	8	2.56	-0.91	15	0.380
	Yes	7.33	9	3.39			

Table 4.2: Shows that recall was significantly higher when a memory strategy was applied in all conditions except the virtual two-room (VW2).

To view the effectiveness of memory strategies in virtual conditions compared to real-world conditions across the word lists, recall frequency was plotted for each word delivered by each speaker (See 4.7). In real-world conditions, the apparent "gap" between the two lines indicates a consistent increase in recall when using a memory strategy with better recall for the majority of words (38/40). This suggests that participants could apply memory strategies effectively throughout real-world conditions. In comparison, recall in the virtual world was less enhanced with memory strategies. Fewer words achieved greater recall (27/40) and much more variation for recall improvement when using a memory strategy. This may indicate an inconstancy in applying a memory strategy in virtual conditions.

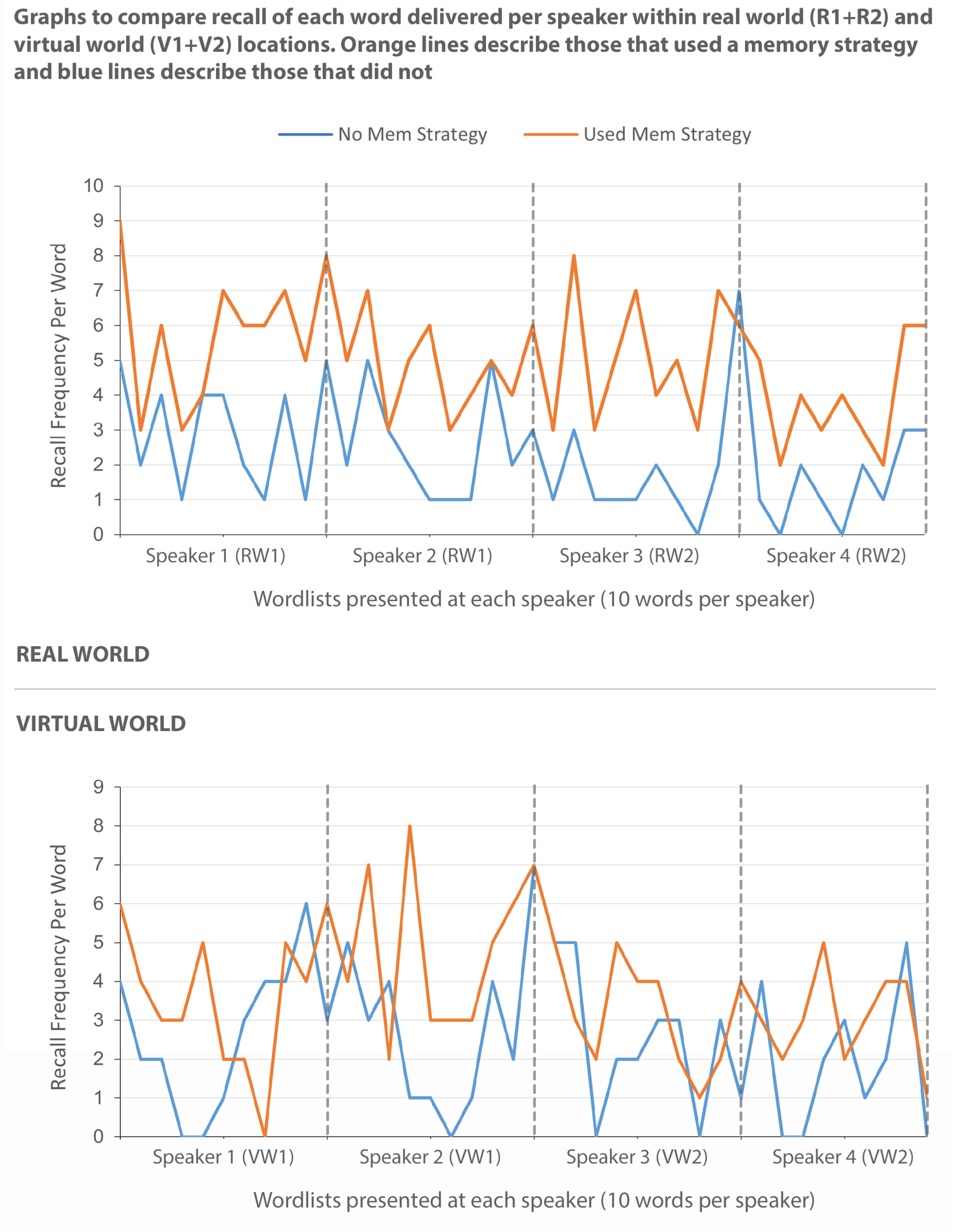

Figure 4.7: The above line graphs show the recall per word for each speaker from those surveyed on whether or not they used a memory strategy (n = 16). The orange line represents those that used a memory strategy and the blue line those that did not. There is much more overlap per word recalled in the virtual conditions between those that used a memory strategy and those that did not. This suggests that the application of memory strategies is more consistent in the real world compared to virtual.

Fig.4.7 compares the real and virtual world recall for each word at each speaker. Primacy and recency effects Murdock Jr (1962) associated with the free recall of lists can be identified in real-world word lists (every 10 words). A primacy effect describes an increase in recall of the first few listed items due to focus of attention (Azizian & Polich, 2007), a chance for rehearsal and minimum interference from reaching working memory capacity on rehearsal processes (Tan & Ward, 2000). A recency effect describes an increase in recall of the last few listed words normally due to strategising the rehearsal process on the last section of a word list and what can currently be recalled for rehearsal from the entire list (Ward, 2002). Importantly, these are established effects that are common to free recall. A reduction in these effects indicates that interference with common rehearsal processes or the participant's attention is divided elsewhere.

To determine the significance of any serial position patterns, the average recall of words was grouped into the beginning, middle and end parts of the word list:

- First 3 words = Beginning
- Middle 4 words = Middle
- End 3 Words = End

If the beginning words in a list are statistically greater than the middle words, this will evidence a significant primacy effect. If the end words are significantly greater than the middle ones, this will suggest a significant recency effect.

For both the virtual and real-world conditions, the initial words of each list were significantly more likely to be recalled, virtual world: $Mdiff = -0.11(11\%), SD = 0.26, t(28) = 2.33, p = .027, 95\%CI[0.01, 0.20]$, Hedges's $G_{a}v = 0.61$, real world: $Mdiff = -0.18(18\%), SD = 0.22, t(28) = 4.42, p = .000, 95\%CI[0.10, 0.26]$, Hedges's $G_{a}v = 0.85$. Although both were statistically significant, the beginning words for real-world conditions were, on average, recalled 7% more than virtual conditions. This difference is statistically significant, $M = 0.8, SD = 0.17, t(28) = 2.59, p = .027, 95\%CI[0.02, 0.15]$, Hedges's $G_{a}v = 0.38$. Although both realities evidence a primacy effect, the real-world condition is statistically stronger. The increase in recall of end words compared to middle, was statistically significant for real-world conditions but not virtual, $Mdiff = -0.11(11\%), SD = 0.21, t(28) = 2.80, p = .009, 95\%CI[0.03, 0.19]$, Hedges's $G_{a}v = 0.54$. This evidences recency effects for the real-world conditions but not the virtual world. These results corroborate that primacy and recency effects are dampened in virtual conditions compared to real-world conditions.

4.4 Discussion

Previous real-world studies suggest that segmenting information between multiple rooms can aid recall (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016). We investigated if similar results could be found within an immersive virtual context and explored the validation of previous results by using a similar methodology in a real-world context. Neither virtual nor real-world contexts replicated previous results. Within a VR context, no significant difference in recall was observed when segmenting information between multiple immersive virtual rooms. In the real-world context, recall from word lists delivered within a single room significantly outperformed all other conditions. Analysis revealed that the type of memory strategy used strongly mediated the free recall results. Reordering words to suit personal associations for individuals produces significantly better recall rates than rehearsal.

Memory strategies can improve word list recall and can supersede the use of incidental environment cues at recall (Smith, 1985). It is, therefore, not surprising to observe their impact on a free recall test. Those that used a memory strategy did not see a benefit to memory strategy application by segmenting information between rooms. This validates previous segmentation work by Smith (1985), which found that those tasked with recalling word lists using a story-based memory strategy did not improve recall by encoding this information between multiple rooms. This highlights that the attention of the individual mediates encoding information. Cognitive resources are finite, so if an individual is focused on internal memory strategies, they will ignore associating information with other external cues, such as the environment. It appears that an external organisation of information does not improve internal memory strategy structures through room locations. However, it should also be noted that there was a low participant number of 16 for the memory strategy analysis. This shows that memory strategies have a large effect, but the nuance of this effect cant be analysed in this study. Using a memory strategy in the real world elicited a clear gain in recall across nearly all words. In the virtual condition, improvements in recall were less consistent. The VW1 condition showed significantly better recall whilst using a memory strategy but not significantly in the VW2 condition. This would suggest some interference when structuring the memory strategy in virtual conditions, with greater interference in the VW2 condition. Additionally, the virtual conditions dampened the primacy and recency effects compared to the real-world conditions. Although using memory strategies will also flatten the standard serial position curve as expected from the free recall of word lists, it is still expected that the primacy effect should be present (Herrmann, Geisler, & Atkinson, 1973; Roediger, 1980). The distraction task will limit the recency effect from any rehearsal, but since there are two word lists in each condition, it is expected that recency effects occur between the first and second word lists. Therefore it would be expected that some evidence of recency effects are present, if not reduced compared to primacy. The dampening of both serial positioning effects and application of memory strategies within the virtual conditions suggests greater difficulty in encoding and recalling in a virtual condition compared to the real world. Through this lens, the significant increase in recall for the RW1 condition can be explained as the easiest environment to apply a memory strategy.

The reduction in memory processes in the virtual world compared to the real world can be explained by the distractions felt by participants within the virtual world. Of the participants who felt some distraction, three themes emerged. The excitement of being within a virtual world, awareness of the VR hardware, and an unsettled feeling towards the VR world representation or from the understanding that their reality has been replaced by a virtual one. These themes were noted to be distracting to the point that they may have interfered with the memorising task. Compared to the real world, this distraction could have reduced the ability to complete the memorising task whether the participant was using a memory strategy or not. These themes account for a reduction in the effectiveness of memory strategy application and support the notion that a single room in the real world had the highest recall due to being the easiest location to apply memory strategies. This appears to be an issue with practice at the experimental conditions. If participants are feeling distracted by feelings of excitement and unsettled by the room representation, then more exposure to these rooms and required activities within may reduce these feelings. Participants had time to acclimatise but did not practice the experimental conditions. Adding an experimental practice round to the tutorial may reduce distraction during trials. The tutorial did appear to be effective at teaching the required interaction techniques as the majority of participants felt they understood the controls after the tutorial and were positive in their comments towards this. Additionally, they felt that they were used to the VR environment, which suggests that they had acclimatised to the VR world in general but not the experimental conditions. This suggests that the tutorial approach is effective at developing interaction understanding and aiding acclimatisation more generally to the VR world.

This study supports previous work in this thesis that segmenting information between two immersive VR rooms does not mediate recall. Additionally, the RW2 condition of this study does not show a significant increase in recall compared to the RW1 condition. Therefore this study did not replicate findings from Pettijohn et al. (2016) and supports the idea that segmenting information between rooms does not affect recall. One explanation is that the voice used for the word lists and the look of the speakers delivering the word lists were the same for all conditions. This could have created a fan effect where the similar details of the speaker's voice and aesthetic between rooms cause all rooms to be accessed in memory on word list recall. The similar speaker voice and aesthetic are causing competition for environmental cues at recall, negating the benefit of segmenting the word lists between rooms. This fan effect may also explain why there is a reduction in primacy and recency effects for virtual conditions compared to real-world conditions. The real-world benefits from much more random detail, even in similar-looking rooms. The virtual conditions had more overlap in room aesthetics which could cause more interference and competition at recall. Additionally, participants would have had instructions from the experimenter when completing the trials and the distraction task. This contact may remind the participants of the real world. This leaves the potential that at times the participant was processing both the virtual and the real world, which could negatively impact encoding processes due to attention being divided between the two realities.

4.5 Conclusion

Building upon previous work that suggests separating information between locations can aid recall, we investigated if segmenting word lists between two rooms could help recall compared to a single room. This segmentation of information was repeated in both the real world and within an immersive virtual reality context. This study suggests that segmenting word lists between real or virtual rooms does not aid recall.

To support this experiment, a virtual reality tool was developed that aided in acclimatising the participants to the virtual reality and help teach the required interaction and navigation paradigm. Participants reported understanding the interaction paradigm and the majority did not feel distracted by the hardware or software. From the reported perceptions of the participants, the virtual reality system does support spatial understanding similar to the real world. However, a dampening of serial position effects and application of memory strategies may suggest a cognitive overhead when using virtual reality experiences compared to the real world.

4.6 Limitations

The interview methodology had limitations. Firstly, the use of closed questions that received a yes or no answer does not enable an empirical analysis of the extent to this agreement. A better approach would be to use a Likert scale questionnaire exploring several questions surrounding each. This would give detail on the extent of the agreement and allow some more meaningful significance testing.

The open, semi-structured questions in the interview require better data capture and a formal coding approach. The current data capture relies on the interviewer to make notes on key points mentioned by participants. This means the interviewer is focusing on questioning, coding, and transcription at the same time. Therefore is more likely to miss information through this capture and also allows interviewer bias to influence the coding of this data. Audio recordings that are transcribed would create a qualitative dataset that could be explored through thematic analysis later through a research team. A thematic analysis ((Hamm, Money, Atwal, & Ghinea, 2019) would enable a variety of lenses to explore this data concerning the questions of the study. These improvements to the interview methodology will aid the "trustworthiness of the study" (Kallio, Pietilä, Johnson, & Kangasniemi, 2016).

The tutorial used to teach controls and acclimatise participants to the virtual world requires an experimental condition practice session. This should set expectations and reduce feelings of excitement or anxiety due to an unknown experience ahead.

Chapter 5

Experiment 03

Walking through virtual doors: Exploring variation in local delivery of information and location change on memory

5.1 Chapter Summary

This chapter describes the methodology, results and analysis of experiment three. Building on experiments one and two, this study investigates if variation in the audio delivery of word lists and visual representation of the speaker mediates recall when the word lists are placed between immersive virtual rooms. The factor of information variation is added to this study to investigate if this can remove the potential fan effect as described in experiment two. This study uses a repeated measures, two-factor design. One factor is whether or not the audio and visual representation of the word lists are the same or varied between word list pairs. The other factor is whether or not these word lists pairs are placed within a single VR room or between two VR rooms. The conditions are as follows:

- V1 = Varied information delivery characteristics of word lists within a single immersive virtual room
- V2 = Varied information delivery characteristics of word lists between two immersive virtual rooms
- S1 = Similar information delivery characteristics of word lists within a single immersive virtual room
- S2 = Similar information delivery characteristics of word lists between two immer-

sive virtual rooms

Experiment three uses the same methodology as experiment two in that participants must listen to two 10-item word lists held within one room or segmented between two rooms. Participants must recall as many words as possible from these lists. Experiment three is preceded by a pilot (n=15) to assess the quality of the VR system used for the main study. Experiment three additionally refines the approach of experiment two by addressing identified limitations. Participants are now given a tutorial for the experimental procedure and a chance to practice this protocol. Contact with the experimenter was identified as a potential immersion-breaking point in the previous experiments. The addition of a narrator now guides participants through the experiment and removes the need for contact with the experimenter during the trials. Additionally, the distraction and recall task is now administered within the virtual world but in a separate neutral VR room. This is the same approach used bySmith et al. (1978). Using a neutral room reduces any impact of context reinstatement (using environmental cues currently around you that were present at the time of encoding to aid retrieval). Not doing so can be seen as a limitation for Pettijohn's Pettijohn et al. (2016) work when segmenting word lists between locations.

The hypothesis for this study is:

H1: Varying the visual and audio attributes of two word lists, will increase recall of these word lists when delivered within separate immersive virtual rooms.

5.2 Method

This study uses an experimental repeated measures design. Two independent variables were identified. Firstly, the number of rooms used to segment the word lists. Secondly, whether or not the audio and visual attributes of the word lists were varied between each or kept similar. These independent variables were treated as factors with two levels each (See Figure.5.1)

The dependent variable was the amount recalled from the word lists for each condition. Using a repeated measure design, the single-room level acted as a control for segmenting information between rooms. The similar information delivery level acted as a control for the information delivery variation factor.

The goal of this study is to have each participant listen to two 10-item word lists for

		ROOM NUMBER	
		One Room	Two rooms
INFO DELIVERY VARIATION	Varied	Varied + 1 Room (V1)	Varied + 2 Rooms (V2)
	Similar	Similar + 1 Room (S1)	Similar + 2 Room2 (S2)

Figure 5.1: Factor diagram highlighting the four conditions for this study

each condition. After a distraction task, they will attempt to verbally recall as many words as they can for that current condition. This recall will then be compared across conditions. Starting position, condition, word list content, and speaker variation will be counterbalanced across participants. With this design, the experiment replicates the base experimental design by Pettijohn et al. (2016), but also allows for comparison of local and global environment details upon recall. Differences from Pettijohn's methodology are outlined as follows. Firstly, participants are asked to listen to each word list twice instead of once. This is to aid the memory of information. Previous experiments in this thesis have shown low recall well below the potential ceiling of information recalled. Secondly, this experiment will use a separate virtual room for recall. An identified limitation of Pettijohn's design (and used in experiment 3 of this thesis) is that participants were instructed to recall in the last room that they encoded information. There is a potential for a context reinstatement effect to occur (Godden & Baddeley, 1975) where participants can use the environmental cues around them to help remember the information encoded within that room. By recalling in a separate environment, the effect of context reinstatement on recall is controlled for. Thirdly, participants are asked to walk around the room for 30 seconds before interacting with the word lists and verbally describe what they see. This is to help acclimate the participant to each room's environmental cues. A simple room scan can lead to a reasonable knowledge of object placement (Helbing, Draschkow, & Võ, 2020) and establish environmental cues.

5.2.1 Materials

This study used Meta Quest 2 (Quest, 2022) hardware to display the virtual reality world. The Quest 2 is an all-in-one head-mounted display (HMD) with a controller for each hand. This device performs "inside-out" tracking, which means the headset calculates the 3D position of the HMD and controllers. Quest 2 does not require any leads to be attached to a PC to run. Each eye receives 1832X1920 pixels with a horizontal field of view (FOV) of 98 degrees and a vertical FOV of 93 degrees. The refresh rate can go as high as 120 Hz, but this simulation's framerate was kept above 70 frames per second to avoid simulation sickness. The Quest 2 controllers performed all interactions within the virtual reality locations.

Unreal Engine 4.26 (Engine, 2004) was used to build all virtual environments. The floor space used for the VR world was 570cm X 630cm, sufficient for participants to explore all environments and rarely see the guardian boundary. The guardian boundary is a wall-sized grid to inform of real-world objects close to the user. As this might remind the participant of the real world, a larger floor space was used to minimise the chance of the guardian boundary appearing.

A set of virtual rooms was developed for participants to acclimatise to the virtual world, gain experience with the navigation and interaction paradigm, understand what is re-

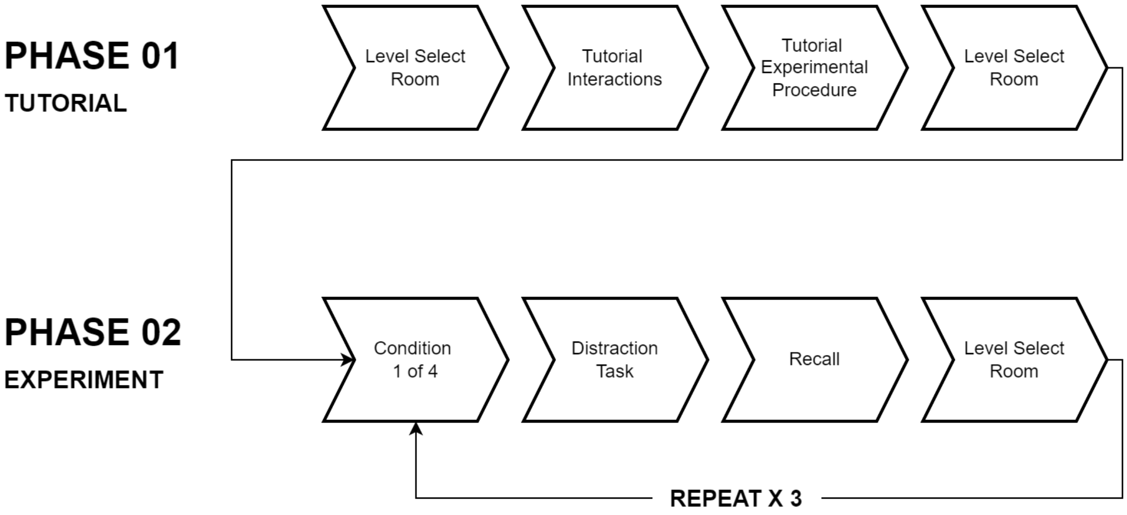
PHASE 01
TUTORIAL
Level Select Room
Tutorial Interactions
Tutorial Experimental Procedure
Level Select Room
PHASE 02
EXPERIMENT
Condition 1 of 4
Distraction Task
Recall
Level Select Room
REPEAT X 3

Figure 5.2: Flow diagram to illustrate the sequence of events required to teach the player the required interaction and navigation paradigms, experimental condition protocol and recall capture. This sequence is split into a tutorial phase and an experimental phase to highlight the point at which the participants are expected to have sufficient knowledge to take part in the experiments.

quired from them during the experimental conditions, and progress through the study. All interactions are done through a single "trigger" button on either controller. The sequence and content of these rooms were designed around an iterative tutorial sequence that leads into the experimental conditions and recall data capture. See Figure.5.2 for the flow of events and outline of goals at each stage. Once the tutorial levels are complete, two aspects of the participant are assumed: They have had enough guidance and practice to understand the navigation and interaction paradigms. The participant understands what is required of the experimental procedure and has practised this.

The participant is returned to the level-select room during the experiment phase. Now when they stand on the level select menu icon it reads "Start Trial n/4". The participant will be taken to one of the four conditions. Each condition will have two, 10-item word lists to listen to twice each. These word lists will be presented as speakers to the participant to interact with. The position of these speakers when placed within a single room or between two rooms are always the same distance apart from each other (2m) (See Figure.5.4). However, the room shape surrounding the speakers and general aesthetic will be different.

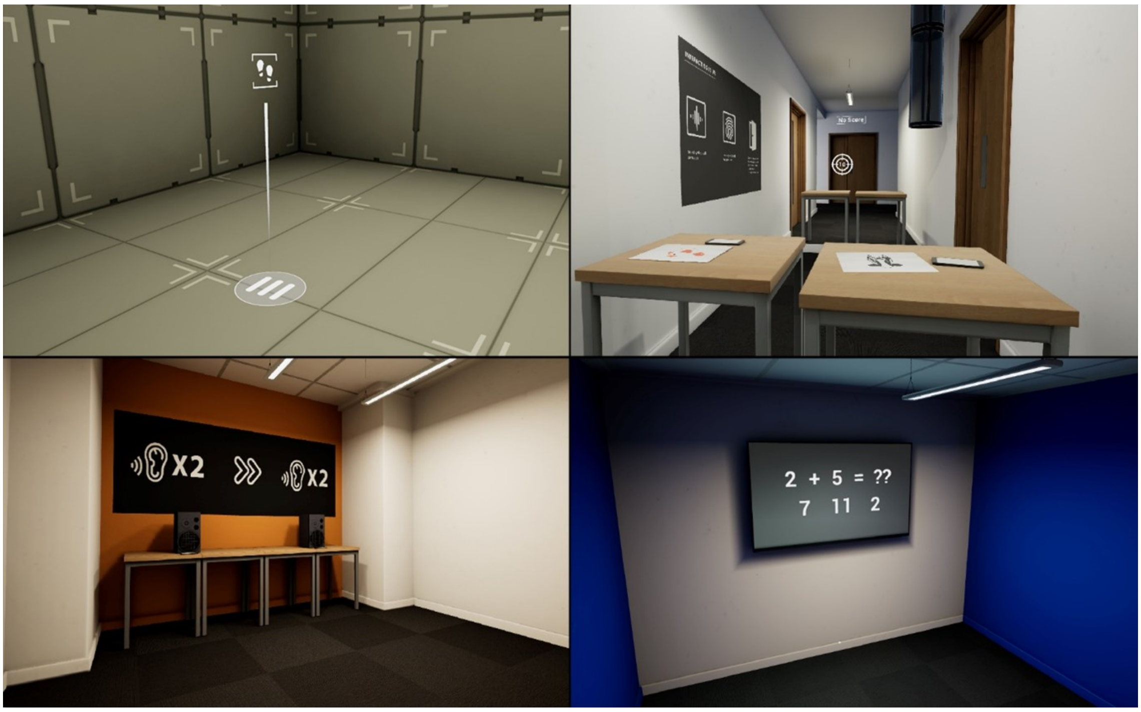

Figure 5.3: Images of the tutorial phase levels. The top-left image is the level select room with the menu activation switch and iconography to walk towards it. The top-right is the interaction training room. The bottom-left is the experimental procedure training room, and the bottom-right is the recall room with maths questions as the distraction task.

The level Select Room: Has three goals. Firstly, this is where the participant transitions to all levels within the VR system. Secondly, it helps teach participants to walk around the virtual area and make basic selections with the controller. It does this by signifying that the user should walk to the "menu" symbol. When reached, a menu appears in front of the user to select their next level. Thirdly, due to the menu symbol being moveable, it is used to position the player before subsequent levels.

Interaction Training Room: Has three main goals. Firstly, it is used to teach interactions with doors and icons through tasks described on a wall poster and directions from the narrator. One task requires the user to find a button behind a door to release balls for a throwing challenge. The challenge is incidental, but the act of repeatedly going through the door aids practice at this motion. This room is also a chance to calibrate the hardware before the experimental conditions. Lastly, this room provides some time (5 minutes) to acclimatise to the VR hardware and virtual worlds.

The Experimental Procedure Training Room: Uses interactions and navigation techniques taught thus far. The role of this room is to explain the procedure and have the user practice the experimental condition protocol. It uses two speakers that deliver word lists different to the main conditions.

The Recall Room: provides a neutral space for a distraction task between encoding and retrieval of word lists and a place to verbally recall the word lists. Users are given a mock trial within this room before the experimental trials begin. The procedure is the same. Users are asked to select the right answer to a set of maths questions for 30 seconds when they load the level. They are then asked to recall as many words as possible from the previous room.

The Experimental Conditions: provide four unique environments for the word lists. These rooms are either one-room or two-room designs. The procedure is the same each time the participant enters a trial room. The narrator informs them to look around the environment for 30 seconds verbally describing it. They are then to approach each speaker twice and memorise the word lists best they can.

For a video walkthrough of the virtual levels, and conditions, refer to "EXP 3" work presented in the supplementary material (Watson, 2023)

Due to the repeated measures design, four distinct environments were created for the four conditions of this study (See Figure.5.4 and 5.5). This was required as a separate event model should be established for each condition. Using the same environment for multiple conditions could create a fan effect because similar environmental cues across trials may

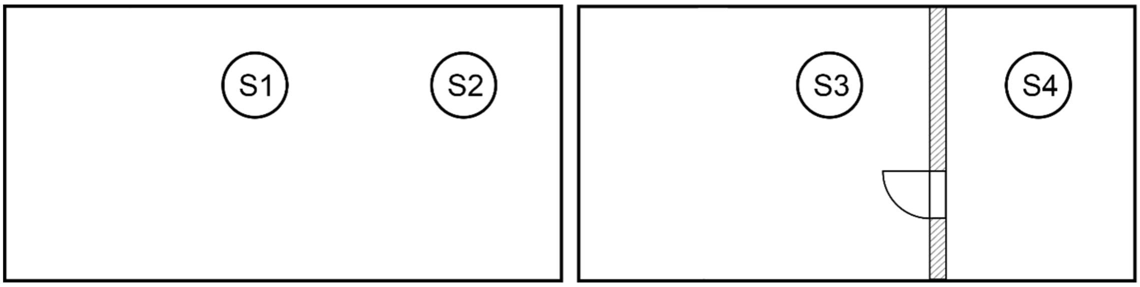

Figure 5.4: These floorplans show the general layout of the experimental conditions. Even if the condition requires a single room or two rooms, the distance between the speaker is kept the same to help standardise the time between each speaker across conditions.

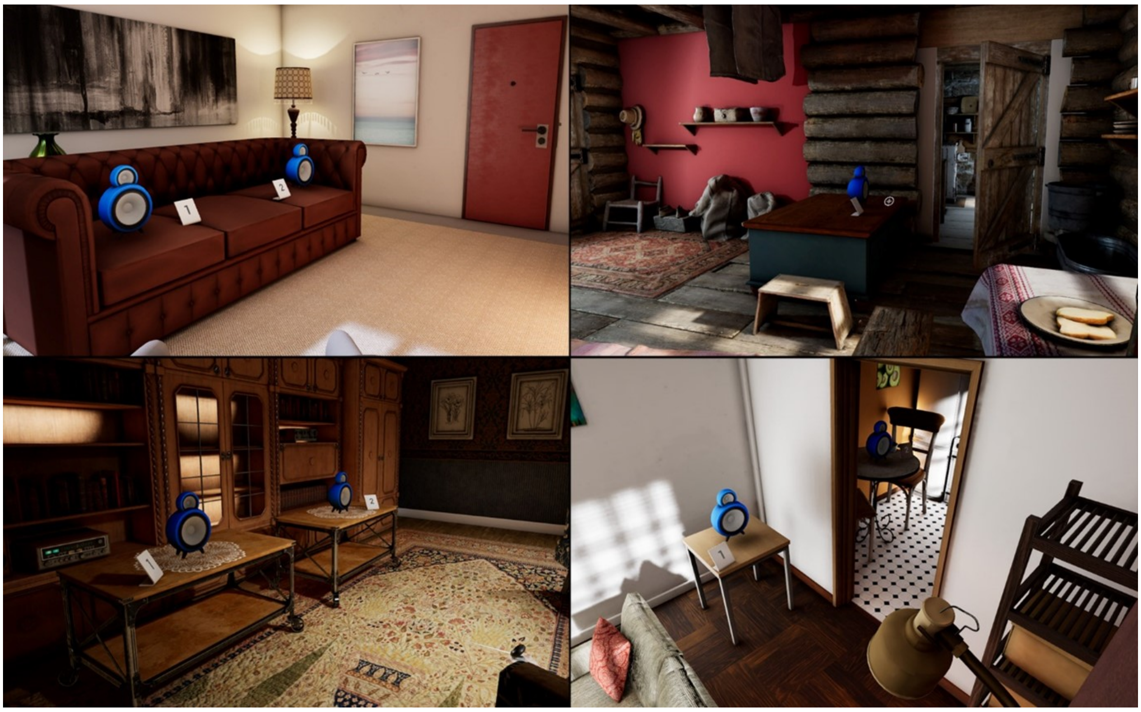

Figure 5.5: These four images show the four environments used for this study. On the left are the two single-room conditions, and on the right are the two-room conditions. Between conditions and rooms within each condition, the aesthetics and lighting are varied to foster the use of distinct environmental cues for each room and condition.

interfere with recall. Therefore, each environment was given a unique aesthetic, and the lighting conditions also varied between environments. For the two-room environments, the lighting was also varied between each room. Additionally, each room was given a distinct context and overarching colour to help segment the between-room environmental cues to foster event model creation for each room.

One factor that establishes the conditions is whether or not there is variation in the visual and audio characteristics of the speakers when delivering word lists. To vary the visual look of the speakers when required, six diverse models were created. Size, shape and colour were altered whilst retaining the general look of a speaker (See Figure.5.6).

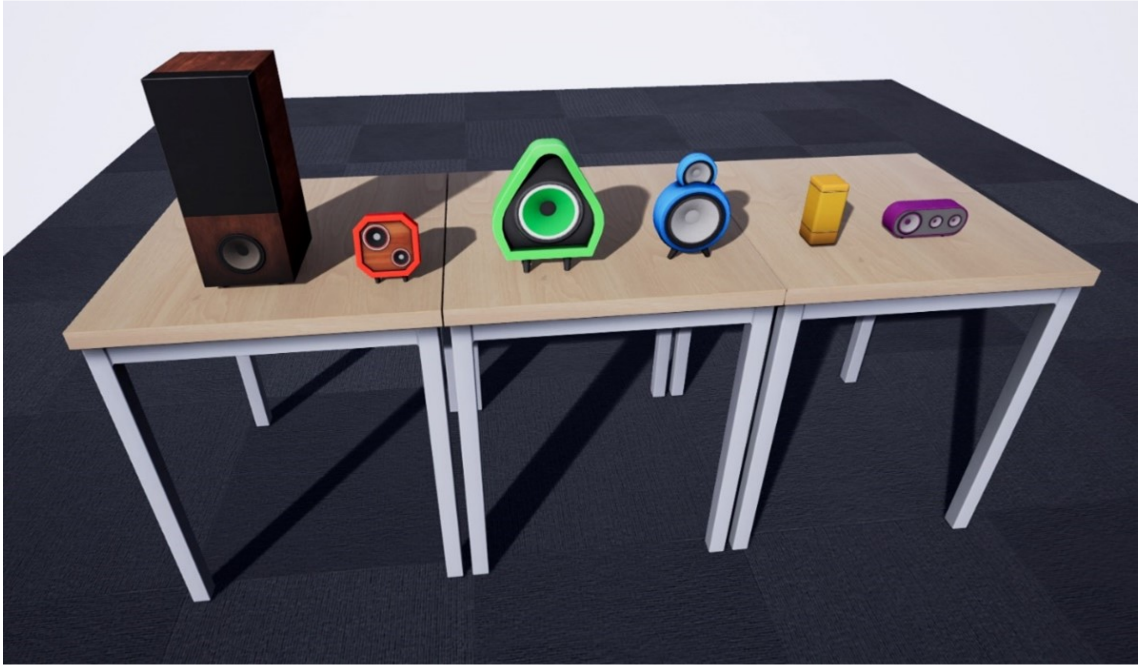

Figure 5.6: Image to showcase the six speakers required to add variation in visual attributes across rooms and conditions

Across the four conditions, eight word lists were used (two word lists for each condition). The audio of these word lists was varied by using six different voices. These voices were generated through Murf-Studio (2022). With this software, a script can be uploaded, and various AI-trained voices can be used to read through the writing. These voice files can then be downloaded and used within the conditions. Three female and three male voices were used. For each gender, two English-sounding and one Canadian-sounding voice was used. Importantly, voices were used that could clearly communicate each word. Voices and word lists were counterbalanced across each condition. For each condition that required variation, one, male and one female voice was used. When participants have gone through all four conditions, they will have been exposed to all voices and all word lists. Overall, six visual and audio deviations of the speakers were required to account for variation between and within conditions.

Eight word lists were required for this study, two for each condition. Each word list was 10 items long. The MRC Psycholinguistic Database was used (Database, 1997) to generate the words. Each word was one syllable, five letters long, and ranged in word frequency from 20 to 103 per million. Words were randomised into eight sets of 10. These word lists were then reviewed to make sure there were no prominent word associations within each list.

Surveys were created using JISC online surveys tool. Participants completed these surveys on a desktop PC within the same laboratory as the VR simulation. Before being exposed to the VR experiment, a survey was administered to the participant to establish demographic information and previous experience in using VR platforms. Post-VR experiment surveys were used to capture data on tutorial efficacy, the usability of the system, sense of presence, simulation sickness, cognitive load, memory strategy approach, and perceptions of voices used in the study.

Simulation Sickness data was captured through the Virtual Reality Simulation Sickness (VRSQ) questionnaire (Kim, Park, Choi, & Choe, 2018). This questionnaire is based on a more widely used Simulation Sickness Questionnaire (SSQ) (Kennedy, Lane, Berbaum, & Lilienthal, 1993) but focuses on more specific VR symptoms of simulation sickness. There was a symptom that participants would feel in this study that was not to do with simulation sickness but was included in the analysis of other simulation sickness questionnaires like SSQ. This was, sweating. In this study, participants will move and interact with a virtual world for 30 minutes. This act will cause sweating, especially when the headset rests upon the participant's face.

The Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann, & Regenbrecht, 2001) was administered to quantify presence. This questionnaire has been used in various other studies, allowing for comparison against other VR experiences. The survey also attempts to distinguish between the spatial and interactive elements of presence. This taxonomy of presence elements is useful when trying to better understand participants' spatial cognition as they explore a variety of environments.

Usability data of the VR experience is also gathered. This study uses a VR experience with a unique navigation and interaction paradigm. It is important, therefore, to evaluate if the paradigms chosen facilitate or hinder the user from achieving the study goals. If poor usability is observed, this would suggest difficult, frustrating or inefficient interactions in the perception of the user. Such an experience could distract from the tasks asked of the participant and, to a degree, invalidate data collected. To assess this, a Simple Usability Survey (SUS) (Brooke, 1996) was modified and administered. The SUS gives an overall score out of 100. Numerical bands within this score can be abstracted to more understandable adjectives (good, poor, excellent etc) (Bangor, Kortum, & Miller, 2009). Additionally, the survey questions can be categorised into "how usable "and "how learnable" the system is (Lewis & Sauro, 2009). The modifications to the original survey are to phrase the language of the software and hardware around VR technologies. Additionally, three extra questions were added to gather more detail on the experience specific to this study:

1. I found the features to be "buggy" (at times faulty, not working as expected).
2. I found the controls/interactions to be "buggy" (at times faulty, not working as expected).
3. I found the menus easy to use.

It should be noted that two questions were omitted from the original SUS survey. This was a mistake. Since the total SUS score is averaged across questions, both the minimum and maximum potential score with these two questions added will be provided to give a realistic range of this score and the average between. An estimation of "How learnable" can still be established from the questions administered. Additionally, when used alongside other data, usability can be interpreted. The questions omitted were:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex

This study uses a tutorial phase to teach and apply interactions/navigation paradigms, experimental procedures, and memory activity. To understand the extent to which participant's felt the tutorial phase prepared them for the experimental conditions, a survey was administered. This survey would show the participant an image of each level and then, through a Likert scale, ask to what extent participants could achieve the aims of that room.

Additionally, participants were asked to view their experience through the lens of cognitive load. Cognitive load refers to the attentional and working memory resources available to engage with an activity and achieve the goals that drive the activity (Van Merrienboer & Sweller, 2005). Cognitive load is often segmented into three types. Intrinsic cognitive refers to the complexity of the task in terms of interactions and subject matter. High complexity or volume will use more cognitive resources. Extraneous load refers to the extent to which material is presented that can distract from any intended learning. Germane load refers to any mental resources required to add information to long-term memory. The ability to add information to existing long-term stores and understanding will depend

Experiment Cognitive Load Survey

Conative Load Lens	Question Asked
Intrinsic Load – Understanding of the Task	I was aware from the start of the experiments that I needed to remember the words form the speakers The instructions during the memory activity were not clear.
Intrinsic Load – Word Memorising Complexity	The words used in the memory activity were very complex. Some of the words used in the memory activity were unknown to me. There were too many words to memorise
Extraneous Load	The virtual world distracted me from focusing on the words presented. The VR hardware distracted me from focusing on the words presented. The VR interactions distracted me from focusing on the words presented.
Overall Cognitive Load	I found the word list activity complicated.
Relative Cognitive Load	Memorising and recalling the words was the most difficult part of this VR experience.

Figure 5.7: Questions asked to explore perceptions of cognitive load during the experimental phase

on the number of mental resources taken up by both the intrinsic and extraneous loads. If intrinsic and extraneous loads fill mental resources, then little to no information will pass to long-term storage. Therefore, it is important to assess if the cognitive load is high when performing an experiment involving memory. If the task is complex and the user is distracted, their ability to recall words in line with normal behaviour is significantly diminished. Additionally, attention towards environmental cues may also diminish. In a VR experiment, it is also important to distinguish if the cognitive load comes from the task asked of the participants or the interactions within the virtual world. Figure.5.7 highlights the questions asked regarding cognitive load and how these questions fit within aspects of cognitive load theory. Responses were in the form of a Likert scale.

Previous studies conducted towards this thesis have found that participants in this experimental design commonly use memory strategies. Using memory strategies to aid recall will significantly mediate what is recalled and will most likely supersede any influence of segmenting information between virtual rooms. Therefore, any strategy used needs to be quantified to help explain the results. To compare the approaches used to memorise the word lists, two survey questions were administered:

1. When memorising the word lists, did you attempt any memory strategies? If so what were these?
2. When memorising the word lists, were you consistent with your approach for each speaker? If not, can you describe how you varied your approach for the speakers?

Data was also collected on the participant's perceptions of the AI-generated voices used in this study. If the voices were perceived as unrealistic, they could be distracting to the task. Additionally, the clarity of words is important as the instructions spoken by the simulated narrator and word lists delivered across all voices needed to be understood. To assess this, a series of questions were asked based on voice clarity, voice distinction between word lists and perception of whether or not the voice was realistic or artificial.

5.2.2 Procedure

This study was run in the same laboratory for all participants. Participants took part in the study, one at a time. After arrival, participants were asked to fill in a consent form and discuss the study before starting. A demographic survey was then completed. Participants were then guided to the middle of the floor space and asked to put on the VR HMD. Once the headset was adjusted for comfort and clarity, the experimenter placed virtual controls in the hands of the participants and informed them that only one button was required for this study and that no other buttons should be pressed. This button is the "trigger" button and is easily pressed by the index finger.

At this point, the participant is in the tutorial phase of the study. Firstly, the participants were asked to stand on the circle in the centre of the Level Select room and select *"Start Simulation"*. The narrator then guided participants to explore the Interaction Training Room. The participants were asked if the audio clarity was sufficient upon selecting the sound objects. If the participant did not find the interactable doorway, they were prompted toward it by the experimenter. After five minutes in this room, the participant was transported to the Experimental Procedure Training room. The simulation narrator guided the participant through the experimental procedure and then asked the participant to listen to both speakers. This guidance is as follows:

Ok, now that you are familiar with the controls and how to interact with this world, I will explain what you need to do in each experimental level.

When a level starts, you will be placed within a room with some speakers that you can interact with. The location of the speakers may be in one room or within multiple rooms. Through these speakers, you will listen to words read out to you. Your task will be to remember these words and complete any challenges asked of you the best that you can. For each level please approach the speakers in numbered order, listen to the information twice, and then move on to the next speaker.

"Please practice this now on both speakers in front of you."

After listening to both speakers twice, the participant is transported to the Recall room. They are instructed to complete the maths questions presented. After 30 seconds, they are asked to *"say out loud as many words as they can remember from the previous room's speakers"* Participants are given one minute to complete this challenge. They are transported back to the Level Select room when this minute has passed. This concludes the tutorial phase of the study.

Once again in the Level Select room, the participant must now find the menu icon on the floor and stand upon it to activate the menu selection for *"Start Trial"*. This will transport the participant to the first condition.

Once transported to the condition trial, the participant is asked to *the next 30 seconds, please explore the environment and vocally describe what you see."* This is to encourage the development of environmental cues and acclimatise to the new room. The participants are then asked: *"When ready, walk in front of, and listen to both speakers twice, in numbered order."*

Once both speakers have been listened to twice, the participant is transported to the Recall room where they must complete a set of maths questions for 30 seconds and then given one minute verbally recall as many words from the previous room's speakers as they can. Once completed, the participant is taken back to the Level Select room. The participant then repeated these steps for the next three conditions.

Once all four conditions were completed, the participant was taken back to the Level Select room and informed by the simulation narrator that the trials were over. With the help of the experimenter, the participant then removed the VR hardware and was invited to complete the post-experiment survey. This survey includes questions on simulation sickness, task difficulty, the efficacy of the tutorial, usability of the virtual system, approach to memorising the word lists, and perceptions of the voices used within the study.

5.2.3 Timings

For each participant, the entire study process lasted 45 - 50 minutes:

- Consent and demographics survey (5 mins)
- Equipping VR hardware and adjustments (2 mins)
- Tutorial Phase of the simulation (10 mins)
- Experimental Phase of the simulation (22 mins)
- Post-experiment surveys (5 - 10 mins)

5.3 Pilot

This pilot received approval from the Research Ethics Integrity Committee, University of Plymouth. A pilot study was conducted to validate the methodology (n=15 Male = 3). Participants were a convenience sample of students studying Psychology degrees at the University of Plymouth. They were recruited through the Psychology Participant Pool. In taking part in this study, students earn credits to gain access to the participation pool for their future experiments. The pilot assessed whether the VR system established immersion, adequately taught the required controls and experimental procedures, and explored the extent to which cognitive load was due to the VR system.

The effects of room number and variation on the recall of word lists were analysed with an ANOVA test. There were no statistically significant effects, with Room F(1,14)=0.06, p=.803; Variation F(1,14)=1.93, p=.186, and the interaction F(1,14)=0.003, p=.955. This was expected from the low sample size of the pilot. The VR system scored low on the VRSQ score (Occ = 10.00, Dis = 8.89, Total 9.44/100) suggesting that the navigation and interaction paradigm does not induce simulation sickness. Presence scores measured through the IPQ were similar to immersion from a VR game experience (Schwind, Knierim, Haas, & Henze, 2019). Realness presence scored below the seven-point middle value of 4.00 (M = 3.97, SD = 0.72) with general presence (5.13, SD = 1.3), spatial presence (M = 4.88, SD = 0.98) and involvement presence (M = 4.80, SD = 0.99) scoring significantly higher. This suggests immersion was established in line with other interactive experiences. Counterbalancing of voice used for information, visual variation between the speakers, and word list used for information delivery were analysed separately as one factor with multiple levels through an ANOVA. There were no statistically significant effects, with Word List F(1,7)=1.13, p=.352; Voice Used F(1,7)=1.01, p=.427, and the Speaker Aesthetic F(1,7)=0.285, p=.958. This suggests no individual word lists, the voice used, or speaker aesthetic held a significant advantage for recall of the word lists.

After analysing responses to the usability, tutorial, and cognitive load questionnaires, participants felt confident interacting with the experimental conditions. They felt they understood what was required of them by the time they took part in the trials. The most cognitively demanding challenge was the memorisation activity. This was not due to a lack of understanding or complexity of the task. This was due to the number of words required to learn between the two speakers. Presence was established, and signs of simulation sickness were very low. Counterbalancing of word lists, voice used, speaker to approach first, and visual look of the speakers did not show any significant signs of influencing recall either. This suggests that the clarity, understanding, and word difficulty across word lists and voices are relatively equal and do not bias recall. AI-generated voices were used to produce the required breadth of voices across all word lists and guide the user through the VR experience. Participants felt that these voices were natural, clear and distinct. Therefore, the audio for this methodology is fit for purpose. These results highlight that the tutorials reached their goals with the majority of participants and that the experimental conditions do explore the separation of information between VR rooms with minimal distraction from VR hardware or VR interaction paradigms.

During this pilot, several participants highlighted being distracted by VR hardware or interactions. The cause of this distraction related to bugs in the system, some further dialogue prompts required for participants to understand that the two-room conditions had an openable door, and a word list was from the experimental condition was duplicated in the tutorial. These were rectified for the main study.

5.4 Results

5.4.1 Participants

This study received approval from the Research Ethics & Integrity Committee, University of Plymouth. 44 (Male = 9) participants were recruited. This was a convenience sample of students studying Psychology degrees at the University of Plymouth. They were recruited through the Psychology Participant Pool. In this study, students earn credits to gain access to the participation pool for their future experiments. On screening, the only requirement was that participants could read, understand and speak English fluently. The majority of participants were aged 18-25 (39/44), with two-thirds claiming they had used a virtual reality platform before (29/44). However, experience with VR is still considered low. Of those that have used a VR platform, only 1 claimed to own one. Additionally, no participants attested to using VR regularly each week, and very few each month. Predominantly, they had used VR "Less than once a year". In response to the context of use, many participants used VR for other experiments, not personal use. In response to the question "Would you consider yourself to be technologically savvy?", participants responded neutral or agree. This indicates experience with standard technologies and some confidence in the ability to explore new technology (See Figure.5.8).

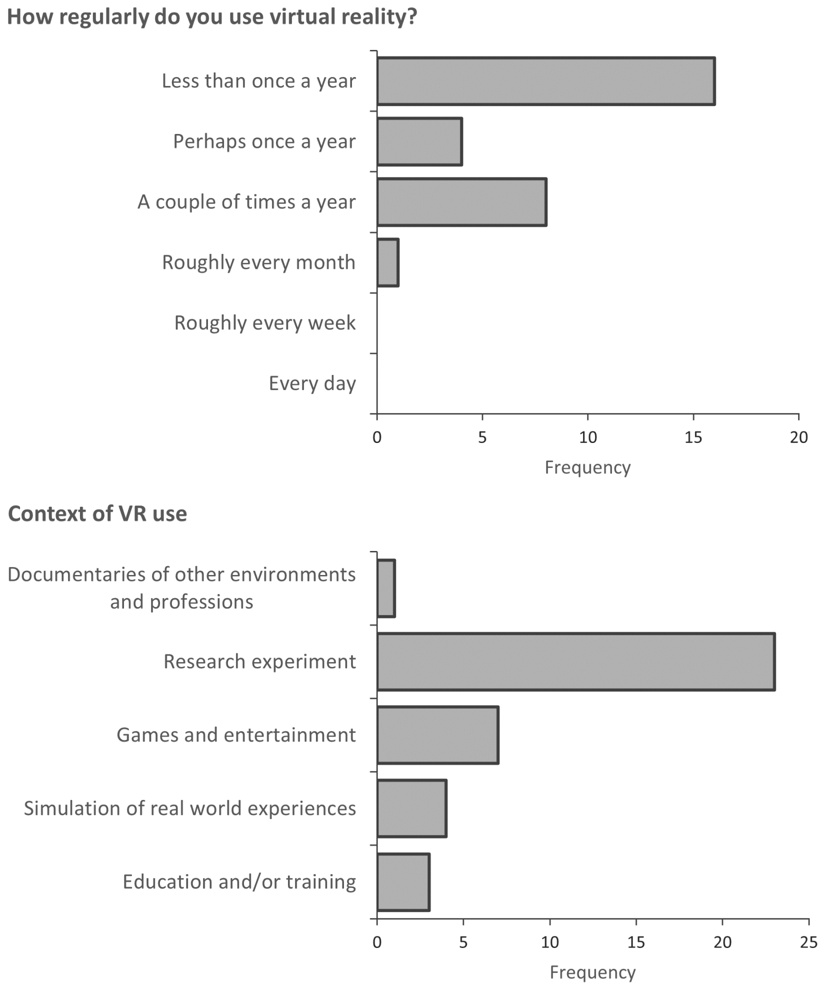

Figure 5.8: Graphs to show regularity and context of VR use. Overall, participants had low experience with virtual reality, confined to other virtual reality experiments.

5.4.2 Recall

For the 44 participants in this study, V1 (M = 6.75/20, SE = 0.689) and V2 (M = 6.96/20, SE = 0.691) conditions observed higher recall than S1 (M = 6.23/20, SE = 0.690) and S2 (M = 6.16/20, SE = 0.558) conditions. Recall data for conditions V1, S1, and V2 were not normally distributed as assessed by Shapiro-Wilk's test (p = < 0.05), so a Wilcoxon signed-rank test was used to compare recall differences for each condition. Statistically, there was no significant median difference in recall between conditions (See Table.5.1).

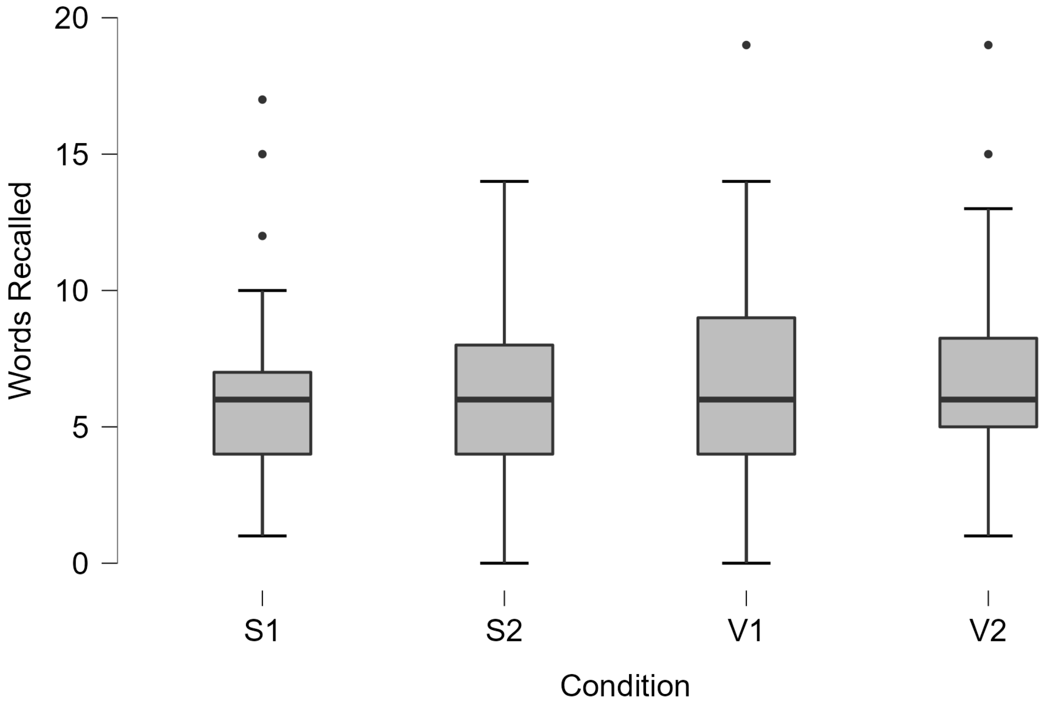

Figure 5.9: Boxplots for the number of words recalled per condition.

Wilcoxon Signed-Rank Test

Pair			n	Statistic	z	p
S1	-	V1	44	295.5	-1.088	.276
S1	-	S2	44	435.5	0.065	.953
S1	-	V2	44	264.5	-1.751	.078
V1	-	S2	44	490.5	0.778	.437
V1	-	V2	44	267.5	-0.232	.822
S2	-	V2	44	263.5	-1.328	.184

Table 5.1: Wilcoxon signed-rank test comparing each condition. No significant difference between the recall for each condition was observed.

To analyse if there were any interactions between the factors of room number and information variation, the recall data was first log-transformed for normal distribution. An

ANOVA was then conducted on this log-transformed data. Factors of room number and variation were analysed with an ANOVA test. There were no statistically significant effects, with Room F(1,43) =0.07, p=.792; Variation F(1,43)=1.533, p=.222, and the interaction F(1,43)=0.669, p=.418.

To explore the effect of participant approach to encoding and recall of the word lists, a post hoc analysis was conducted comparing those that used a memory strategy against those that did not. 26/44 participants reported using a memory strategy to aid the recall task. Of those that reported using a memory strategy, 11 created a story from the words, six created associations to categorise words, one linked the words to actions, five associated words to environmental objects, and three linked the words to own experiences. Of those that used a memory strategy, 24/26 reported being consistent with their approach across all trials. The effect of memory strategy was significant, F(1,42)=27.78, p=<.001. Memory strategy interactions with Room Number (F(1,42)=0.737, p=0.395), Variation (F(1,42)=1.418, p=.24), and both Room Number X Variation (F(1,42)=0.3.911, p=.055), were not significant.

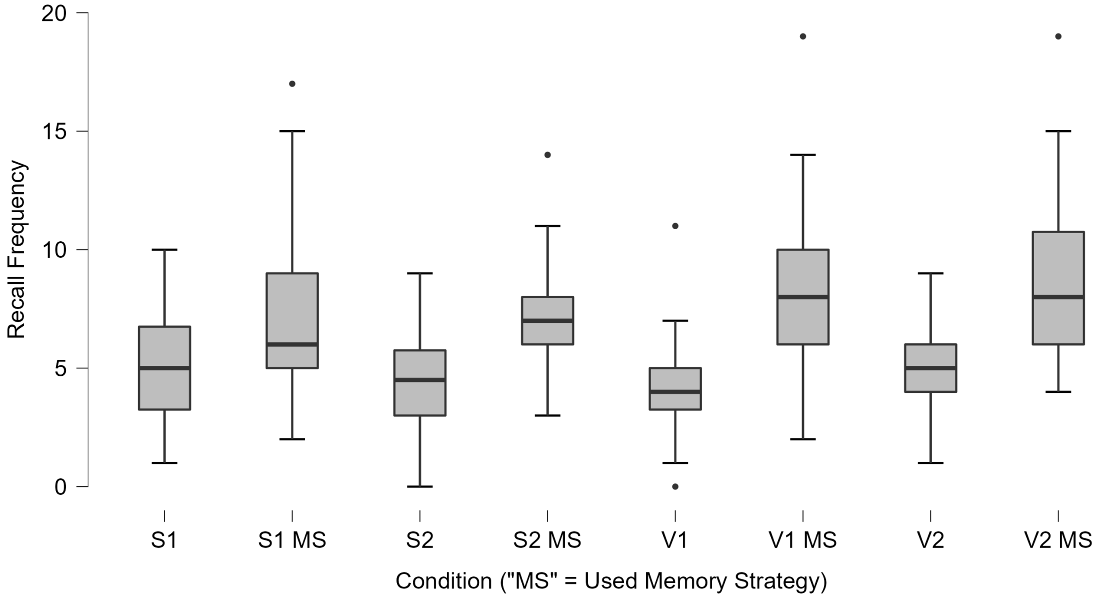

Figure 5.10: Boxplots that compares recall for those that used a memory strategy against those that did not for each condition.

Due to the significant impact on recall from those that used a memory strategy having much higher maximum recall scores, the data was explored further using only those that had used a memory strategy (n = 26). Participant recall data were non-normally distributed and so log-transformed for analysis through an ANOVA. A significant effect of Variation was observed, F = (1/25) 5.576, p = .026. Factors of Room Number (F(1,25)=0.0.834, p=.370) and interaction between Room Number and Variation (F(1,25)=0.454, p=.507) were not significant. Paired sampled t-tests were performed between each condition. The S1 condition, which used a single room and similar visual and audio characteristics for the information delivery, showed significantly lower recall than when variation was present for information delivery.

To analyse the extent of the variation significance, paired samples t-tests were performed across all conditions. Information variation was not a main effect across all conditions, but conditions V1 and V2 showed a significant recall increase compared to the S1 condition. The S1 (M=7.04, SD =3.52) condition observed significantly (t(25) = -2.32, p .029) lower recall compared to V1 (M=8.39, SD =3.51) and significantly (t(25) = 2.67, p .013) less recall compared to V2 (M=8.42, SD =3.52).

Friedman Test

Factor	Chi-Squared	df	p	w
Voice Used	5.332	7	0.62	0.017
Word List Used	9.391	7	0.226	0.03
Speaker Aesthetic	6.819	7	0.448	0.022

Table 5.2: Friedman tests across the counterbalanced voices used, speaker aesthetic, and word list. No significant impact on the recall was observed, suggesting that the word list used, voice heard, or form of the speaker did not significantly influence the recall of the word lists across the conditions.

To test if differences between the voices used, aesthetics of the speakers and word lists biased the results, the recall data for each was gathered and compared. A Friedman test was run on each due to the non-normal distribution of the data (See Table.5.2). Within each of these categories, there were no overall significant differences. This suggests that the word list used, voice heard, or form of the speaker did not significantly influence the recall of the word lists across the conditions. However, looking at the distribution graphs, voice M3 did perform significantly worse than voices F3, M1, and M2 (see Figure.5.11. This has some import on the study. The same male and female voices were used for the Similar Information conditions across all participants but counterbalanced for order of appearance. A different set of voices were used for the Varied Information conditions but counterbalanced for order of appearance. Since M3 observed significantly lower recall than its counterbalanced F3 voice, then the voice itself could have caused lower recall rates. This should be taken into consideration with previous results showing an impact of variation on those that have used a memory strategy. Since this impact is due to an underperforming S1 condition where the voice used for word list delivery is kept the same, then it is probable that the relative under-performance of S1 is caused by the M3 voice used in these conditions. This argument is strengthened by the F3 voice used in the S1 conditions having relatively similar recall rates to the other voices.

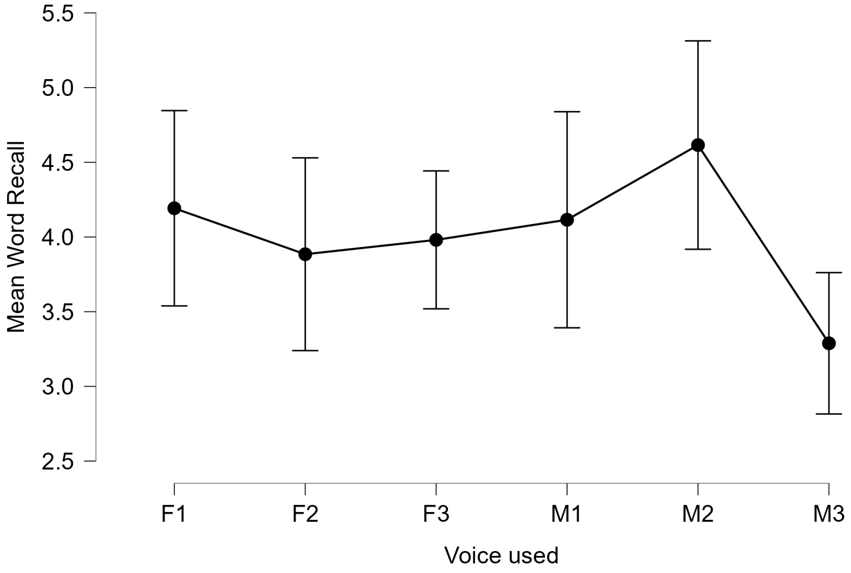

Figure 5.11: Graph that compares mean recall for each voice used during word lists delivery for those that used a memory strategy. Voice M3 had significantly less recall than voices F3, M1, and M2

5.4.3 VR System

Post-exposure questionnaires were used to analyse the VR system from the perspective of usability, simulation sickness, presence, cognitive load, navigation and interaction efficacy, and experimental protocol understanding. Many of these categories were explored through five-point Likert scale surveys. To test significance, a one-sample t-test was administered to the average of these scores against the scale's mid-point. A significant average above the midpoint will suggest some agreement with the statement, and a significant average below will suggest disagreement.

Responses to the SUS survey suggest generally positive usability (See Table.5.3). Taking into account the two questions not administered, the minimum SUS score from the available questions averages 62.50, with a maximum value of 82.50 and an average between this range of 72.50. The minimum score can be described as "Ok", with the maximum and average described as "good" (Bangor, Kortum, & Miller, 2009). Since the average between the range is "good", it is more likely that this is the most accurate response if all questions were administered. There is a significant positive response for this system

Usability Survey Results

Usability Category	Mid-Point	Mean ± SD	t-value	df	p
SUS Questions[a]					
VR controls were easy to use	3.00	4.39 ± 0.62	14.88	43	< .001
Support of a technical person[b]	3.00	3.75 ± 0.99	5.02	43	< .001
Functions well integrated	3.00	4.39 ± 0.58	15.87	43	< .001
Too much inconsistency[b]	3.00	4.30 ± 0.77	11.23	43	< .001
People would learn this quickly	3.00	4.27 ± 0.76	11.13	43	< .001
VR controls were very cumbersome[b]	3.00	3.73 ± 1.07	4.53	43	< .001
Confident using the VR controls	3.00	4.11 ± 0.72	10.23	43	< .001
Learn many things before use[b]	3.00	4.07 ± 1.11	6.40	43	< .001
Learnable	3.00	3.91 ± 0.87	6.92	43	< .001
Project Specific Questions					
Features were "buggy"	3.00	4.52 ± 0.66	15.20	43	< .001
controls/interactions were "buggy"	3.00	4.32 ± 1.01	8.69	43	< .001
Menus were easy to use.	3.00	4.57 ± 0.50	20.76	43	< .001

[a] Questions have been simplified in table for formatting purposes

[b] Values reversed to gather positive perception from disagreement to negative statement

Table 5.3: SUS scores show sentiment towards the usability questions and how learnable the software is. P-values values suggest that the average response is significantly away from the mid-point of 3.00 in the Likert scale SUS survey. In this case. all responses are positive as they are significantly above the mid-point of 3.00. However, the high standard deviation of learnable questions suggests a neutral to strongly agree response in this area.

being "learnable". However, taking into account the standard deviation, this response could fluctuate from neutral to strong agreement, which suggests a positive but wide variation in the strength of the positivity. This is reflected in questions surrounding the learning of the system. Although confidence in using and learning the controls is positive, there is variation in whether or not technical support is required. Functions and features are viewed as well integrated without significant perception of bugs.

The VR system provides a tutorial for interaction and navigation and the experimental procedure required through trials. When surveyed (Table.5.4), participants agreed they could easily interact with the required icons and doors. Additionally, they felt they did not require more time in the tutorial area to refine these skills. Although significantly positive, there is wider variation in agreement as to whether participants could interact automatically (not think about their interactions) post-tutorial. The standard deviation shows a range between neutral/disagreement to strongly agree.

Participants were surveyed on their experience with the tasks of the experimental trials (Table.5.5). During the experimental conditions, participants felt they knew what they

Tutorial Efficacy and Room Goals

Questions	Mean[a] ± SD	t	df	p
- Interaction Training Room -				
By the end of the interaction tutorial room, I felt that I could easily open and close the virtual doors.	4.68 ± 0.52	21.53	43	< .001
By the end of the interaction tutorial room, I felt that I could easily interact with highlighted icons to generate sounds and spawn balls.	4.71 ± 0.70	16.12	43	< .001
I feel that I needed more time in this tutorial room to comfortably learn how to interact with this virtual reality world.	1.71 ± 0.98	8.78	43	< .001
By the end of the interaction tutorial room, I did not have to think about how to interact with the VR world, it happened automatically.	3.77 ± 1.12	4.59	43	< .001
- Experimental Procedure Training Room -				
I found the explanation of the experiment procedure easy to understand.	4.64 ± 0.49	22.31	43	< .001
Upon leaving the procedure explanation room, I felt as though I understood what was required of me in the experimental conditions.	4.50 ± 0.51	19.67	43	< .001
Upon leaving the procedure explanation room, I felt that I needed more time to prepare myself for the experimental conditions.	1.64 ± 0.61	14.75	43	< .001

[a] Mid-point of the five-point Likert scale is 3.00. 1 = Strongly Disagree, 5 = Strongly Agree

Table 5.4: Mid-point value of the five-point Likert scale is 3.00. Significant mean values above 3.00 suggest agreement with the statement and below disagree. Results suggest that the tutorial rooms effectively taught the navigation and interaction paradigm. However, although significantly positive, there is wide variation in how automatic participants' interactions were with the virtual environments after the tutorial.

were doing. Each environment was perceived as distinct. There was significant agreement that the two-room environments felt like different rooms. The standard deviation suggests some range in this agreement from neutral to agree/strongly agree. Whether or not the participants were distracted by the VR hardware did not reach significance for either agreement or disagreement. The standard deviation and non-significant results suggest a range between agree and disagree. Looking closer at the data, 23 disagreed/strongly disagreed that the VR hardware was distracting, 14 agreed/strongly agreed that it was, and 7 were neutral. In the recall room, participants knew how to proceed with distraction and recall tasks. There was a non-significant neutral response to whether or not the participants found it awkward to verbalise their recall. However, there was significant

Experimental Conditions Experience

Questions	Mean^{a} ± SD	t	df	p
- Experimental Conditions Rooms -				
When I went through a doorway, it felt as though I was entering a new room.	4.46 ± 0.66	14.55	43	< .001
Interacting with an icon to start a word list was cognitively demanding.	2.05 ± 1.06	-6.00	43	< .001
I was unsure of what I was doing in the experimental conditions.	1.64 ± 0.75	-12.06	43	< .001
The rooms felt very similar.	1.89 ± 0.95	-7.81	43	< .001
I found it easy to navigate the experimental conditions.	4.41 ± 0.50	18.79	43	< .001
Opening and closing a door was cognitively demanding.	1.86 ± 0.93	-8.11	43	< .001
I was distracted by the hardware (controllers, headset etc) during the experimental conditions.	2.64 ± 1.22	-1.97	43	0.055
Any level with two rooms felt like two separate rooms.	4.11 ± 1.13	6.57	43	< .001
Each level felt like a different room	4.43 ± 0.79	12.03	43	< .001
- Recall Room -				
I found it difficult to select the answer for the maths question task	1.80 ± 1.09	-7.33	43	< .001
I did not know what I was doing with the maths question task	1.32 ± 0.60	-18.55	43	< .001
After the tutorial, I knew what I had to do in this room	4.41 ± 1.04	8.98	43	< .001
I found it awkward to verbalise words from the previous room	3.09 ± 1.07	0.56	43	0.578
During the experiment I found this room confusing	1.50 ± 0.59	-16.85	43	< .001
I soon got used to verbalising words within this room	4.25 ± 0.72	11.53	43	< .001

[a] Mid-point of the five-point Likert scale is 3.00. 1 = Strongly Disagree, 5 = Strongly Agree

Table 5.5: Mid-point value of the five-point Likert scale is 3.00 and therefore significant mean values above 3.00 suggest agreement with the statement and below disagree. Results suggest that verbally recalling the word lists felt initially "awkward", but that participants soon got used to this process.

agreement that they soon got used to verbalising words in this room.

Cognitive load perception during experimental trials was surveyed (See Table.5.6). Memorising the recalling words were perceived as the most challenging element of the experimental trials. Through the lens of cognitive load, participants significantly agreed that the memory activity was not complex, instructions were clear, and they were not

Experimental Activity Cognitive Load Perception

Questions	Mean^a ± SD	t	df	p
I was aware from the start of the experiments that I needed to remember the words form the speakers	4.05 ± 1.24	5.60	43	< .001
The words used in the memory activity were very complex.	2.11 ± 0.75	-7.80	43	< .001
Some of the words used in the memory activity were unknown to me.	2.00 ± 1.18	-5.62	43	< .001
The instructions during the memory activity were not clear.	1.36 ± 0.49	-22.31	43	< .001
The virtual world distracted me from focusing on the words presented.	2.80 ± 1.15	-1.18	43	0.246
I found the word list activity complicated.	2.09 ± 1.05	-5.73	43	< .001
There were too many words to memorise.	3.73 ± 1.02	4.73	43	< .001
The VR hardware distracted me from focusing on the words presented.	2.07 ± 1.17	-5.29	43	< .001
The VR interactions distracted me from focusing on the words presented.	2.11 ± 1.15	-5.13	43	< .001
Memorising and recalling the words was the most difficult part of this VR experience.	4.64 ± 0.65	16.69	43	< .001

^{a}Mid-point of the five-point Likert scale is 3.00. 1 = Strongly Disagree, 5 = Strongly Agree

Table 5.6: Mid-point value of the five-point Likert scale is 3.00. Significant mean values above 3.00 suggest agreement with the statement and below disagree. Results suggest that cognitive load was focused on the memorising activity. However, there is some variation on whether or not the words presented were known to the participants and disagreement on if the VR worlds distracted participants from focusing on the words presented.

distracted by the VR hardware or interactions. However, all these responses also showed standard deviations ranging from disagree/neutral to agree/strongly agree. This suggests a bias to agree with some feeling distracted or found the trials complex. When asked if the virtual world distracted them from focusing on the words, the response was neutral and not significantly away from the mid-point value. A closer look at the data shows 20 disagreed/strongly disagreed, 8 were neutral, and 16 agreed/strongly agreed with this statement.

Participants were surveyed on their perception of the AI-generated voices used for narration and word list delivery (Table.5.7). There was significant agreement that the narrator's voice was clear and easy to understand. Although significant agreement on this voice seemed natural, there was a considerable variation from agree to disagree. There was significant agreement that the voices delivering words were clear and that participants

Voice Perception

Questions	Mean^a ± SD	t	df	p
The narrator's voice was clear and easy to understand	4.41 ± 0.82	11.45	43	< .001
The voices delivering the words were clear and easy to understand	4.05 ± 1.01	6.86	43	< .001
The narrator's voices was annoying	1.82 ± 0.87	-9.01	43	< .001
The narrator's voice seemed natural	3.52 ± 1.09	3.19	43	0.003
The voices delivering words generally seemed natural	3.82 ± 0.92	5.89	43	< .001
I could easily tell when different voices were being used between speakers	4.21 ± 1.03	7.80	43	< .001
All the speaker voices sounded the same	1.73 ± 1.09	-7.77	43	< .001
The narrator's voice seemed artificial	2.64 ± 1.01	-2.38	43	0.022
The voices delivering words seemed artificial	2.61 ± 1.08	-2.37	43	0.023

^{a}Mid-point of the five-point Likert scale is 3.00. 1 = Strongly Disagree, 5 = Strongly Agree

Table 5.7: Mid-point value of the five-point Likert scale is 3.00. Significant mean values above 3.00 suggest agreement with the statement and below disagree. Results suggest that voices were clear and understood. Although a significant p-value was reached for the perception that voices were natural, there is some variation in this opinion.

could tell these voices were different between speakers. Again there was a significant disagreement that these voices were artificial but some variation in this opinion from agree to strongly disagree.

The VR system scored low on the VRSQ score (Occ = 19.27, Dis = 11.50, Total 15.38/100), with 3/44 reporting some nausea after the trials. In response to the IPQ realness presence scored below the seven-point middle value of 4.00 (M = 3.93, SD = 0.92) with general presence (5.523, SD = 1.42), spatial presence (M = 5.33, SD = 0.89) and involvement presence (M = 4.61, SD = 1.06) scoring significantly higher.

5.5 Discussion

Previous real-world studies suggest that segmenting information between multiple rooms can aid recall (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016). We investigated if similar results could be found within an immersive virtual context and explored if variation in the audio and visual details of information delivery could aid this effect. No significant difference in recall was observed when segmenting information between rooms or when adding variation to the visual and audio attributes of information delivery. Post hoc analysis revealed that memory strategies strongly increased the free recall results compared to rehearsal strategies.

This study closely followed the methodology for segmenting information between rooms as established by Pettijohn et al. (2016). It also refined the VR implementation compared to previous studies developed in this thesis by controlling for potential fan effect, context reinstatement, removing contact with the experimenter during the trials, and providing an improved tutorial for both interactions and experiment procedure. This study, therefore, suggests that within VR, the incidental processing of the environment is not a strong enough organisational framework to enhance recall of word lists by segmenting their location between rooms.

This is not to say that within real-world contexts, segmentation of information between rooms may still bare some positive aid to recall, but in an immersive virtual context, this is unlikely. However, this study does not cover all elements linked to improved recall of word lists through room segmentation. Previous work has identified that increasing the time between word lists (walking between locations or waiting outside) may aid free recall Smith (1982). Knowledge of events may increase if there is a time lag (10 minutes) before the free recall task is administered (Flores, Bailey, Eisenberg, & Zacks, 2017). This study and Pettijohn et al. (2016) cannot account for this factor and may be a key component to establishing memory hooks in different rooms. Additionally, this study cannot generalise beyond a cohort of psychology students. This sample was a convenience sample and also matched the type of cohort used by Pettijohn et al. (2016). Due to the knowledge base of Psychology students, there may be a bias toward using memory strategies compared to other population samples. One improvement in design compared to Pettijohn et al. (2016) is that memory strategy use has been recorded and evaluated. The methodology used leaves room for memory strategies to be utilised. Using word lists and a defined challenge of "remember the words the best that you can" may encourage the use of memory strategies. With such a strong effect and a high number of participants using such an approach, this methodology requires some form of memory strategy analysis to frame the results.

Presence was established with low levels of simulation sickness, which suggests the VR platform and experience would not have significantly distracted participants through feelings of nausea or breaking immersion. The usability of the VR system is good. This suggests that participants could achieve the goals of the tasks asked of them. When surveyed, participants also had confidence in performing the required interaction and navigation. Although generally positive usability was observed, participants would likely feel confident with technical support. This suggests that with a researcher to guide them, the participants felt they could efficiently learn how to use the system. Participants also perceived to have enough understanding of the experimental trials by the time they took part. These results suggest some success in the tutorial element of the VR system to train the participants in the VR system and experimental conditions.

However, some points of distraction were highlighted through the variation of participant responses. Although participants were significantly positive in their ability to interact automatically with the VR system, there was variation in this question. More practice time is likely required to achieve automaticity of interaction across the cohort. This might have implications on recall if many participants required more cognitive resources to interact and navigate the VR system during the memorising activity. Likewise, participants felt the controls were not cognitively demanding, but there was still some variation. Additionally, although participants felt comfortable with the control and navigation, they were neutral on whether or not the VR hardware or environments distracted them. This suggests there is a split between those that got to grips well with the VR system and those that did not. Longer acclimatisation in each experimental condition may lessen this distraction. These results highlight a potential impact on recall from lacking automaticity of interactions and distraction from the VR experience. However, the majority felt confident in taking on the tasks asked of them, which suggests the VR system has made good strides in tutorial development but requires more work to be effective for a broader cohort. Therefore the usability should not have interfered with the experimental trials significantly. This also raises an interesting line of inquiry with novel complex systems used in research; how can we quickly get participants up to speed with the controls, navigation, and understanding with and without a researcher present?

The voices used for the word list memory activity were reported to be clear, understandable, and distinct. However, there does seem to be a drop in performance due to the M3 voice used in the Similar Information conditions. This can explain the relative drop in recall performance for the S1 condition observed for those that used a memory strategy and highlights the need for better screening of voices to be used when memory activities are present.

5.6 Conclusion

This study provides some evidence that segmenting word lists between virtual rooms does not aid recall. It also suggests that variation in audio and visual attributes of the word lists delivery does not aid the recall when information is placed within the same room or between two rooms. However, the use of memory strategy was a main effect. More nuanced factors of segmenting the word lists and varying the information delivery were overshadowed.

The tutorial was effective at transferring the knowledge needed to take part in the experimental conditions, acclimatising participants, and giving confidence to the actions of participants within the VR world.

The methodology of this study closely follows previous real word work that observed enhanced recall from segmenting word lists between rooms. Due to the main effect of memory strategy, this study suggests that memory strategies can easily navigate the participant task and therefore needs to be controlled for in future work.

5.7 Limitations

The SUS questionnaire was administered with two questions omitted. A range was therefore given on the potential minimum, maximum and average between. This should be comparable to other studies with this limitation taken into account. However, all questions should be used for more accurate comparability between the scale being used on similar types of systems.

Conclusion

In this thesis, I investigated two themes. Firstly, **can segmenting wordlists between two virtual rooms be used as an environmental design approach for VR training?** The idea was inspired by studies that have found that placing information between real-world environments can aid recall. VR platforms afford spatial cognition similar to the real world whilst enabling navigation between various locations. If there is a strong tie between location and recall of experiences, these attributes could be used to guide environment design that supports the memory of this content. Secondly, **what are the challenges for replicating real-world experiments within VR?** How can we limit the impact of novel VR hardware and experiences so that participants can understand the experimental procedure, learn VR interactions, and acclimatise to the VR world? Although commercially available, VR hardware is not in everyone's house. This means standards of interaction are not commonly practised. Additionally, any good practice is still being explored by developers and content creators as the hardware evolves. One challenge when using a technology that is in development and generally unpracticed by a population is how to ensure all users have a consistent ability to achieve the goals of the experience. Understanding how to onboard users successfully to a VR world without inhibiting an individual's ability to take on the required challenges of that experience is not only key for limiting confounds in research but also helps the efficacy of VR training and adoption across industries.

To explore these proposed themes, this thesis asks the following questions:

THEME 1: Can we utilise the segmentation of information between locations as an environment design approach to aid memory recall for VR training?

R1.1 What is the effect of segmenting information between immersive virtual locations on recall when individuals are not directed on how to encode or retrieve this information?

R1.2 What are the impacts of the visual and audio characteristics of the virtual location on recall when information is segmented between locations?

R1.3 How will individuals approach the encoding and retrieval of information when no specific strategy is required?

THEME 2: What are the challenges for replicating real-world experiments within VR?

R2.1 What is the impact of an immersive VR platform on cognitive processes which support the encoding and retrieval of information?

R2.2 How can we efficiently onboard users into a VR experience so they have sufficient knowledge of navigation and interaction paradigms before experimental conditions?

Can we utilise the segmentation of information between locations as an environment design approach to aid memory recall for VR training?

R1.1 What is the effect of segmenting information between immersive virtual locations on recall when individuals are not directed on how to encode or retrieve this information?

Understanding if segmenting information between virtual locations can aid recall is paramount as to whether this approach can be used to guide the design of environments for VR training. This line of enquiry follows from real-world studies that have found an increase in recall of word lists when segmenting this information between multiple rooms. Experiment one to three replicated a previous methodology from event cognition work (Pettijohn, Thompson, Tamplin, Krawietz, & Radvansky, 2016) where subjects would walk between two ten-item wordlists either within a single room or segmented between two rooms. Across three studies, no significant increase in recall was observed by segmenting word lists between immersive virtual locations. Nor did varying the visual or audio characteristics of the word lists mediate recall. Consistently, the most significant effect on recall was memory strategy use, which was not mediated by room number or information variation.

Further work is needed to explore the use of information segmentation as a design approach for VR. The number of rooms used to segment information may mediate an effect to aid recall. Experiments in this thesis compared two rooms against one. However, Smith (1982) found greater recall of word lists when this information was encoded in four rooms compared to two. Pettijohn et al. (2016) found an increase in recall of a short narrative when two location changes were present in the text compared to one. This may highlight that there is a balance between the volume of information delivered and the amount of segmentation. In both Smith (1982), Pettijohn et al. (2016) experiments, the extra segmentation did not add more words when increasing segmentation. It was the same information across more locations. Other work has observed a strong effect of context for word recall when one word is tied to one context (video background, location etc.). Smith & Manzano (2010) found in a real-world study that one word associated with one video background observed significantly higher recall than 15 words associated with one background. Logie & Donaldson (2021) found that within virtual environments, while moving through 10 rooms, associating four words to each observed significantly better recall than having the same 40 words associated with one room. Logie & Donaldson (2021) suggests that keeping information within working memory limits per context is part of the increase in recall, allowing cognitive resources to move exposed content into longer-term storage. This may suggest that the organising factor of segmenting information between event models is mediated by the amount of content required to learn. Similarly, location-based memory strategies (method of loci) that associate one word or name to a specific location within a room (imagined or virtual) have shown an increase in recall (Legge, Madan, Ng, & Caplan, 2012). This work highlights how effectively associating smaller units of information to specific location landmarks can be for recall, especially when an individual is trained to leverage this method. However, associating single units of information to small contexts may not be a practical approach to designing virtual training environments. It may require user training and limit the information individuals can be exposed to. However, it does raise an important question for future research in segmenting information between locations. What is the balance between information delivered and segmentation between locations for effective recall? Can this be effectively utilised without users' training in a mnemonic strategy? Additionally, if information retention through any form of segmentation is significantly mediated by working memory, this could be used as a design approach for learning paths. Information could be steadily increased per segmentation as the learner gains knowledge. Thus information per segmentation based on previous knowledge could be a design variable required for effective encoding when furthering knowledge from fundamentals to advanced concepts.

Although experiments one to three did not find an enhancement to recall by segmenting word lists between immersive, virtual rooms, This is not to say that in real-world conditions, segmenting word lists between rooms can enhance recall. Through two studies, Pettijohn et al. (2016) demonstrated an increase in recall of word lists when segmented across rooms and windows within a desktop display. Smith (1982) was able to also observe an increase in recall of word lists when segmenting word lists between rooms. However, he could not replicate these findings to statistical significance in future work (Smith, 1984, 1985). Methodologies between Smith and Pettijohn were similar, except that Smith's work had some time between list delivery, and rooms were not next to each other. This inconsistency in previous work and the variation of approach to memorising word lists noted in this thesis suggests some disparity in how participants use location to organise memory of word lists. Neurological observations suggest that location is integral to recalling past experiences (Robin, Buchsbaum, & Moscovitch, 2018). When individuals use location to help recall events, they increase the speed of recall and accuracy for that event (Hebscher, Levine, & Gilboa, 2018). However, not everyone will use location as a cue. The benefit of segmenting information between locations may be individual differences or practice recalling events. If this is the case, participants should undergo a series of tests to identify cognitive ability relating to episodic memory. For example, Sargent et al. (2013) tested participant working memory, laboratory episodic memory, executive function, processing speed, and general knowledge to compare against the event segmentation ability of participants. Segmentation ability was analysed as the degree to which an individual participant's ability to recognise discreet events within videos compared to the norm of the cohort. Segmentation ability was found to predict event memory independently from other psychometric tests (working memory etc.). Therefore, Sargent et al. (2013) concludes that segmentation ability is important to memory and a cognitive mechanism that has variance within a population, much like working memory. Therefore, not evaluating the participant's cognitive mechanisms towards episodic memory is a limitation of this thesis. The ability to segment events may dictate who would benefit from separating information between locations. Additionally, due to immersive VR being able to generate all manner of content, an exciting line of work would be to use VR to supply content that efficiently analyses an individual's cognitive abilities within the context of VR.

R1.2 What are the impacts of the visual and audio characteristics of the virtual location on recall when information is segmented between locations?

A key construct of an event model is the location in which an event takes place. The creation of an event model for each location is theorised to help organise experiences for better recall of information within. Therefore, immersive virtual rooms used to segment information should feel like distinct rooms to help form a distinct event model for each. Colour, shape, form, and purpose can help describe a room but may also influence how significantly we can recall this room. This suggests three interesting questions. Did participants perceive the rooms in these studies as separate rooms? Did any room provide an advantage over another in completing the required memory tasks? Does variation in room characteristics aid recall when segmenting information between rooms?

For experiments one to three, participants were asked if moving within the two-room conditions felt as though they were separate rooms. The vast majority of responses felt as though they were. This suggests similar cognitive processes in understanding a transition between rooms have been established within the immersive VR world. One participant in experiment two highlighted that this transition between rooms was less significant than in the real world. However, others were assertive in their impression of transitioning between rooms. Given the well-known affordance or spatial understanding when using immersive VR platforms, it is likely that those who were engaged with the experience did perceive two rooms in the two-room conditions. This suggests that an event model was established for each virtual room. It is also possible that the context of being within a VR reality was strong, which could create an overriding context that merges these event models in some form. This might make it more challenging to keep these event models distinct. This could be a driving factor for not observing any effect of segmentation across experiment one to three. Exploring the impact of multiple realities on cognition would be an interesting line of enquiry and requires more comparisons between real and virtual worlds. Although experiment two did compare the virtual and real worlds, the effect of memory strategy was strong and confounded the comparison of realities.

A variety of environments were used in experiment one to three. In experiment one, the concept of an art gallery was used to enable distinct rooms with various colours and adorning imagery. In experiment two, the concept of a basic teaching room was used to match real-world conditions. In experiment thee, four distinct environments in terms of theme, colour, shape and purpose were developed. Throughout these varied environments, no one room appeared to have an advantage on recall compared to the others within an experiment. In experiment three, one voice type scored lower, possibly due to a lack of understanding when delivering words. However, in terms of visuals, no speaker appeared to have an advantage for recall, suggesting that the colours and themes used did not significantly mediate results. However, experiment three highlighted that participants were split on whether or not the virtual worlds distracted them from the memory task. This may mean that the worlds developed were too "interesting" and required more time to acclimatise and explore to reduce distraction during memory trials.

Experiment three directly compared recall when information was delivered through varied visual and audio characteristics (different look and voice of the speaker delivering word lists) against the same characteristics. The aim was to evaluate if variation in local (speaker delivering words) and global (room) environment details is required for a segmentation effect to be observed. This experiment did not observe a significant recall change when varying local and global characteristics. This suggests that when designing spoken content for VR, there may be no benefit to recall in using different voices and visual representations of this content.

R1.3 How will individuals approach the encoding and retrieval of information when no specific strategy is required?

One assumption when using multiple rooms to segment information delivery is that individuals would not need to consciously apply a strategy to encode and retrieve this content. The segmentation would create event models that help chunk and act as memory hooks for the segmented content. Therefore, the segmentation encourages a structure that could be broadly utilised without extra conscious effort from an individual. The potential is improved recall using automatic cognitive processes and limited addition to cognitive load. However, consistently across experiments one to three, half of the participants used their memory strategies to aid word list recall. In experiment one, memory strategy had no significant effect on recall. For experiments two and three, using a memory strategy was the significant factor in improving recall. The difference in the effect of memory strategy could be due to the cohort's ability to apply them. Given the strength of the effect in experiments two and three, it would be expected that the same effect would be significant in experiment one. This could be explained by participant variation, that participants in experiment one were not proficient in using memory strategies. Another reason could be that there needed to be acclimatisation to the VR world. So the novelty and adjustment to the new reality impeded the conscious effort to apply a memory strategy effectively. The consistent use of memory strategies poses two crucial questions. Firstly, is this the use of memory strategies a confound for exploring information segmentation between immersive virtual rooms? Secondly, is this approach ecologically valid or an artefact of the experiential methodology?

The use of memory strategies will be a confound for exploring the segmentation of information between immersive rooms. If participants use memory strategies, they will develop internal structures and cues to associate the list of words when encoding. This means they will not use external cues to associate this information. The incidental processing of an environment would be considered an external cue and therefore be ignored by those utilising a memory strategy as attention is focused on internal representations. This makes using memory strategies a confounding variable in this methodology, as those using a memory strategy are not processing external cues. This highlights a limitation of the methodology developed by Pettijohn et al. (2016). Without accounting for memory strategy approach, it is not possible to know if results are due to environment cues aiding organisation or variation in approach to memorising the word lists. Due to limited cognitive resources, participants are likely to organise information in a structure they are either well practised with or is the easiest path to a memory hook. Therefore, when presented with a memory task, they choose to use a strategy if they know one that fits the task. It is unlikely that those using a memory strategy for wordlists would use the same approach for general learning. Creating abstract stories and word associations for a list of words is beneficial for a wordlist memory task but perhaps not helpful when attempting to encode larger bodies of knowledge from various sources (textbooks, videos etc.). Therefore, the memory task methodology used in experiments one to three encourages using memory strategies that are not generally used in educational and training environments.

Discouraging the use of memory strategies should be a focus of future work exploring information segmentation. One improvement to future methodologies would be to alter the content, so it is more difficult to apply a memory strategy. For example, use an action sequence like the recovery position, or deliver a micro-teach. This content would be more ecologically valid in education and training but also require much more development time. Word lists are pragmatic for an experimenter to administer. However, the lack of context within random word lists may bias behaviours towards associative tactics in the form of memory strategies. Rating words might seem like a plausible alternative to the goal of memorising them. For example, rating the words in terms of "pleasantness" would enable the processing of a word without formulating strategies. However, when Smith (1985) used this approach for multiple room segmentation of information, this condition did not observe any difference in recall. Again this could be due to internal associations with the words overshadowing the cues of the environment. Using various information presented virtually is an interesting avenue for segmentation between contexts. Using a between-subject design instead of repeated measures may be advantageous for a memory study. In a between-subject design, a surprise recall test can be used, and the goal for the participants during exposure to the information could be focused on learning, not memory. This may engender behaviours more ecologically valid for learning and storing information and not biased strategies that are effective just for the current trial.

What are the challenges for replicating real-world experiments within VR?

R2.1 What is the impact of an immersive VR platform on cognitive processes which support the encoding and retrieval of information?

VR affords similar cognitive responses to real-world interactions. It can be a fantastic platform to pragmatically replicate real and imagined scenarios to empower research across many domains. Individuals can have their audio and visual senses fully immersed in a virtual replication of the real world. Interactions can be similar, allowing physical walking and simple abstractions of hand interactions through controllers. Furthermore, this is commercially available today. However, concerning the nuance of cognition, it is important to understand to what extent VR facilitate the exact cognitive responses as experienced in the real world. This question is also essential when considering the difference between virtual platforms that undergo significant development each year. Suppose the aim is to share observations between real-world and VR studies. When observations are recorded, how can we be confident that the user's state within one VR platform is comparable to another or the real world? What are the key differences we must consider when using knowledge gained from real-world studies and applying these principles to VR, and visa versa?

Experiments one to three can help answer the extent to which VR impacts cognitive processes when individuals participate in a memory activity. In these experiments, participants were probed on their perception of how distracted from the tasks they felt they were due to the VR hardware or experienced virtual reality. Across these experiments, the majority felt focused on the memory activity. This highlights how modern VR platforms can effectively establish presence and navigation paradigms like walking do not induce simulation sickness. The experimental conditions required the participants to feel they were navigating between different locations. Throughout the studies, it was reported that subjects felt they were passing between rooms, and each area was distinct. This suggests that spatial cognition of the environment was established. Additionally to the platform, a tutorial was developed and iterated upon through experiments one to three to give participants knowledge and practice of the navigation and interaction paradigm and an acclimatisation period. In experiment three, participants perceived themselves to have confidence in their interactions, understanding of the controls, and being able to give good attention to the memory task. However, some points of distraction were highlighted through the variation of participant responses. These pointed to participants feeling that although they understood the required interactions, they were not automatic by the experimental trial. Additionally, the participants were split on whether the VR hardware or world was distracting. Therefore a noticeable section of the cohort could have used more practice and acclimatisation before the experimental trials. Experiment two did observe a reduction in serial position patterns for the virtual, two-room condition compared to real-world conditions, and some evidence suggests that memory strategies were not as easily applied in the virtual conditions. However, the sample size was reduced for the memory strategy analysis, so the fluctuation in a strong effect like memory strategy could be confounded by participant variation. Similarly, McFadyen et al. (2021) attempted to replicate the location updating effect where there is a reduction in recall for information associated with a previous location. Across several experiments in VR, no significant location updating effect on forgetting was observed. However, in these experiments, the rooms were purposefully kept the same, so a fan effect could interfere with memory processes. Similarly, in experiment two of this thesis, there could have been a fan effect from the similar speakers, voices and similar aesthetics between rooms. It may suggest that, to some extent, participants view the whole VR experience as one context due to being part of an experiment and a relatively short excursion to a different reality. This may merge the nuance of different rooms into the same context at recall and cause interference at retrieval. Additionally, there may not have been enough time to adjust to each room when encoding information, causing some cognitive overhead that again interferes with encoding processes. If there is a cognitive overhead to using VR, then lowering the load of the memory activity and increasing acclimatisation time may reduce this overhead. This suggests that exploring the cognitive overhead of an experimental activity is as much about the activity's complexity and the participant's preparedness. Therefore, a helpful comparison between studies would be metrics established at a pilot stage or through a main study as to how much overhead is due to the VR system. This will be more informative when investigating cognitive phenomena with small to medium effect sizes, as these could be confounded by overhead in the VR system.

R2.2 How can we efficiently onboard users into a VR experience so they have sufficient knowledge of navigation and interaction paradigms before experimental conditions?

Developing effective approaches to onboarding users to a VR system will reduce confounds caused by the system and reduce time spent on training participants to take part in any virtual activity. Not only will this lead to more apparent observations of phenomena within experiments, but it will also reduce the strain on participants to take part in these studies. Asking too much of an individual within a given period may fatigue the participant during trials reducing the clarity of observations and ecological validity outside of a fatigued state. Experiment one to three iterated on an immersive VR tool that needed to implement experimental conditions, train participants in experimental procedures, train participants in the navigation and interaction paradigms, and acclimatise participants to the VR hardware and software. By experiment three, the approach to this training had been significantly iterated and showed some good practice for VR onboarding.

A computer game tutorial approach was used to train users in the interaction and navigation paradigm. One interaction concept was approached at one time, practised and then used as the base for the next concept. Therefore, building on what was already known whilst practising previous knowledge gained. Interaction and navigation were taught before the experimental procedure. It would only be possible to practice interacting and to navigate around experimental procedure by first learning these basics. The experimental procedure also enabled more practice of the interaction and navigation techniques. Together, these tutorials acted as an acclimatisation period. Experimenter contact would reduce quickly once the hardware was adjusted and any participant queries resolved. This was to maintain immersion for the participants by not being reminded of the observing researcher in the real world. Contact was unnecessary as an in-world narrator guided the participants through the levels. Thereby the experiment trials, experimenter contact was mostly removed. This approach was largely successful whilst highlighting areas of development in VR tutorials for research.

Teaching interaction, navigation paradigms, and the experimental protocol was the most successful part. Participants felt confident in their tasks and how to achieve them. However, the final tutorial still required the experimenter to watch and give direct guidance on how well participants opened virtual doors. An improvement would be to separate each interaction into a single level and develop an intuitive way to assess the efficacy of such movements by the system. This in itself would be an interesting line of work. How to assess if a basic VR movement has been completed successfully through self-reporting and system analysis. Acclimatisation is an area that requires more improvement. Not all participants felt they had automaticity of controls. Some participants mentioned distraction from the VR hardware and virtual worlds. More time within the VR system may improve participant perception of their own efficacy within a VR space. However, this would also significantly increase each participant's time to complete the experiment. If done in one setting, this reduces participant energy for the trials and efficiency for the researcher to complete their required trials. This may be fine if the trial does not require much time. However, it may be better practice for VR experiments to employ some pre-training hours or days before the trials. This could be a practice session that consolidates movement and experimental information. Then a primer experience could be used before the trial starts. This would give more time for training to accommodate those that take longer to acclimatise to VR platforms and learn the controls.

Acclimatising and teaching users the interaction and navigation paradigms is an interesting line of enquiry. The onboarding metrics can be compared between studies to evaluate how well participants could automatically interact and navigate the virtual world. Experiment one to three uses subjective opinions from participants to evaluate this. Using subjective perceptions, physiological responses, and task performance to develop such metrics would be a constructive step forwards. These could then be used to evaluate the tutorial design to inform best practices for research and training.

Future Work

This thesis explored if the spatial attributes of a VR platform could combine with principles of event cognition theory to help inform the design of virtual environments that support memory recall. To enable these studies, a VR tool was developed that helped to teach navigation and interaction paradigms, experimental procedures, and implement the trials.

Although the use of event cognition theory and methodology did not observe improvements in memory recall, the experiments in this thesis suggest a line of interesting future work to clarify these findings.

THEME 1: Suggested future work to explore information segmentation between locations as an environment design approach to aid memory recall for VR training

- Memory strategies need to be controlled through the use of different content, tasks or between-subject designs.
- Participant segmentation ability and cognitive ability (working memory etc.) should be analysed against recall to establish if there is a correlation between individual differences and final recall.
- Balance of information delivered and segmentation should be explored by dividing the same information across several segmented contexts (two rooms vs three rooms, for example).
- Variation in audio and visual characteristics of the word delivery did not mediate recall across small word lists. However, this needs to be explored across larger bodies of information to know when a change in information delivery characteristics may aid the internal organisation of knowledge and subsequent retention.
- Balance of information delivered between each step or location to be adjusted based on previous knowledge of a learner. Can keeping the amount of information within individual working memory thresholds aid retention?
- Time between information may also be a key factor in consolidating information. This should be varied between information delivery points to assess if information segmentation can aid recall depending on the time interval.

THEME 2: Suggested future work to help onboard participants to immersive VR experiences

The VR tool effectively replicates real-world rooms and delivers experimental trials whilst supporting the training of required knowledge and skills. However, such a tool requires initial experimenter input to help guide participants. This can reduce the ability to use a tool remotely if needed and introduce potential training inconsistencies between participants. The training and acclimatisation for virtual reality experiments are essential variables to consider. Especially when exploring ideas that may transfer to the real world or have smaller effect sizes.

This suggests a line of interesting work concerning the VR platform:

- Explore self-reporting and machine learning approaches to analyse mastery of the interaction and navigation techniques required for the experimental conditions. This would lead to creating metrics that analyse if individuals have attained proficiency with interaction and navigation paradigms. Other metrics to consider are cognitive load and distraction as a means to assess acclimatisation within an environment.

- Finding efficient but effective methods to train participants in the VR controls. Explore tutorial methodologies, including challenge tutorials, to compare against self-reported confidence and task efficacy. Potentials include pre-training sessions that allow users to get accustomed to entering new environments and take on multiple challenges within a navigation and interaction paradigm. This approach also considers the consolidation time of information and the extent to which a primer is needed before experimental trials to bring back learnt lessons from pre-training sessions.

Appendices

Appendix A

Supplementary Material

Experimental forms, surveys, data, and video walkthroughs can be found in the **"Supplementary Material.ZIP"** file, (Watson, 2023)

This file will cover:

Preliminary Study

- Surveys and Forms
- Data Logs
- Video walk-through of the Recovery Position application

EXP 1

- Wordlists used for the conditions
- Data Log
- Grouping methodology
- Video walk-through of the VR conditions

EXP 2

- Surveys and forms
- Wordlists used for the conditions
- Data Log

- Counterbalancing methodology
- Video walk-through of the VR conditions

EXP 3

- Surveys and forms
- Data Log
- Counterbalancing methodology
- Video walk-through of the VR conditions

Bibliography

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, *20*, 1–11.

Ashford, D., Bennett, S. J., & Davids, K. (2006). Observational modeling effects for movement dynamics and movement outcome measures across differing task constraints: A meta-analysis. *Journal of motor behavior*, *38*(3), 185–205.

Azizian, A., & Polich, J. (2007). Evidence for attentional gradient in the serial position memory curve from event-related potentials. *Journal of cognitive neuroscience*, *19*(12), 2071–2081.

Bandura, A., & Walters, R. H. (1977). *Social learning theory*, vol. 1. Englewood cliffs Prentice Hall.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114–123.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology*, *49*(2), 229–240.

Boutin, A., Fries, U., Panzer, S., Shea, C. H., & Blandin, Y. (2010). Role of action observation and action in sequence learning and coding. *Acta psychologica*, *135*(2), 240–251.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of verbal Learning and verbal Behavior*, *8*(3), 323–343.

Brooke, J. (1996). Sus: a "quick and dirty'usability. *Usability evaluation in industry*, *189*(3).

Brooks, B. M. (1999). The specificity of memory enhancement during interaction with a virtual environment. *Memory*, *7*(1), 65–78.

Brooks, D. C. (2011). Space matters: The impact of formal learning environments on student learning. *British Journal of Educational Technology*, *42*(5), 719–726.

Brooks, J. O., Goodenough, R. R., Crisler, M. C., Klein, N. D., Alley, R. L., Koon, B. L., Logan Jr, W. C., Ogle, J. H., Tyrrell, R. A., & Wills, R. F. (2010). Simulator sickness during driving simulation studies. *Accident analysis and prevention*, *42*(3), 788–796.

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends in cognitive sciences*, *10*(12), 551–557.

Choi, H.-H., Van Merriënboer, J. J., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educational Psychology Review*, *26*(2), 225–244.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159.

Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3-d virtual environments? *British Journal of Educational Technology*, *41*(1), 10–32.

Database, M. P. (1997). Mrc psycholinguistic database (dict interface).
URL `https://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm`

Davis, S., Nesbitt, K., & Nalivaiko, E. (2014). A systematic review of cybersickness. In *Proceedings of the 2014 conference on interactive entertainment*, (pp. 1–9).

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of experimental psychology*, *54*(3), 180.

Developer, O. (2022). Performance and optimization | oculus developers.
URL `https://developer.oculus.com/documentation/unity/unity-perf/`

Diersch, N., & Wolbers, T. (2019). The potential of virtual reality for spatial navigation research across the adult lifespan. *Journal of Experimental Biology*, *222*(Suppl_1), jeb187252.

Dzulkifli, M. A., & Mustafar, M. F. (2013). The influence of colour on memory performance: A review. *The Malaysian journal of medical sciences: MJMS*, *20*(2), 3.

Ekstrom, A. D., & Isham, E. A. (2017). Human spatial navigation: Representations across dimensions and scales. *Current opinion in behavioral sciences*, *17*, 84–89.

Elliott, D., Grierson, L. E., Hayes, S. J., & Lyons, J. (2011). Action representations in perception, motor control and learning: implications for medical education. *Medical education*, *45*(2), 119–131.

Elliott, P., & Hawthorne, G. (2005). Imputing missing repeated measures data: how should we proceed? *Australian & New Zealand Journal of Psychiatry*, *39*(7), 575–582.

Engine, U. (2004). Unreal engine | the most powerful real-time 3d creation platform.
URL `https://www.unrealengine.com/en-US/`

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nature neuroscience*, *20*(11), 1504.

Fery, Y.-A., & Ponserre, S. (2001). Enhancing the control of force in putting by video game training. *Ergonomics*, *44*(12), 1025–1037.

Fitts, P. M., & Posner, M. I. (1967). Human performance.

Flores, S., Bailey, H. R., Eisenberg, M. L., & Zacks, J. M. (2017). Event segmentation improves event memory up to one month later. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(8), 1183.

Furlough, C. S., & Gillan, D. J. (2018). Mental models: structural differences and the role of experience. *Journal of Cognitive Engineering and Decision Making*, *12*(4), 269–287.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of verbal learning and verbal behavior*, *5*(4), 351–360.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in cognitive sciences*, *5*(6), 236–243.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology*, *66*(3), 325–331.

Gold, D. A., Zacks, J. M., & Flores, S. (2017). Effects of cues to event segmentation on subsequent memory. *Cognitive research: principles and implications*, *2*(1), 1–15.

Google (2022). Google cardboard – google vr.
URL `https://arvr.google.com/cardboard/`

Hamm, J., Money, A. G., Atwal, A., & Ghinea, G. (2019). Mobile three-dimensional visualisation technologies for clinician-led fall prevention assessments. *Health informatics journal*, *25*(3), 788–810.

Hebscher, M., Levine, B., & Gilboa, A. (2018). The precuneus and hippocampus contribute to individual differences in the unfolding of spatial representations during episodic autobiographical memory. *Neuropsychologia*, *110*, 123–133.

Helbing, J., Draschkow, D., & Võ, M. L.-H. (2020). Search superiority: Goal-directed attentional allocation creates more reliable incidental identity and location memory than explicit encoding in naturalistic virtual environments. *Cognition*, *196*, 104147.

Herrmann, D., Geisler, F., & Atkinson, R. (1973). The serial position function for lists learned by a narrative-story mnemonic. *Bulletin of the Psychonomic Society*, *2*(6), 377–378.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and instruction*, *17*(6), 722–738.

Huttner, J.-P., Qian, Z., & Robra-Bissantz, S. (2019). A virtual memory palace and the user's awareness of the method of loci.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS biology*, *3*(3), e79.

ISO, M. (2018). Ergonomics of human-system interaction–part 11: Usability: Definitions and concepts.
URL `https://www.iso.org/standard/63500.html`

ISS (2016). Recovery position.
URL `https://play.google.com/store/apps/details?id=iss.recovery&amp;gl=GB&amp;pli=1`

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *Neuroimage*, *31*(1), 429–439.

Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education*, *106*, 150–165.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. 6. Harvard University Press.

Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of advanced nursing*, *72*(12), 2954–2965.

Kelly, J. W., & McNamara, T. P. (2010). Reference frames during the acquisition and development of spatial memories. *Cognition*, *116*(3), 409–420.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, *3*(3), 203–220.

Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied ergonomics*, *69*, 66–73.

Kimura, K., Reichert, J. F., Olson, A., Pouya, O. R., Wang, X., Moussavi, Z., & Kelly, D. M. (2017). Orientation in virtual reality does not fully measure up to the real-world. *Scientific reports*, *7*(1), 1–8.

Krause, D., & Kobow, S. (2013). Effects of model orientation on the visuomotor imitation of arm movements: The role of mental rotation. *Human movement science*, *32*(2), 314–327.

Krokos, E., Plaisant, C., & Varshney, A. (2019). Virtual memory palaces: immersion aids recall. *Virtual Reality*, *23*(1), 1–15.

Kuhbandner, C., & Pekrun, R. (2013). Joint effects of emotion and color on memory. *Emotion*, *13*(3), 375.

Küller, R., Mikellides, B., & Janssens, J. (2009). Color, arousal, and performance—a comparison of three experiments. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, *34*(2), 141–152.

Lamers, M. H., & Lanen, M. (2021). Changing between virtual reality and real-world adversely affects memory recall accuracy. *Frontiers in Virtual Reality*, *2*, 16.

Legge, E. L., Madan, C. R., Ng, E. T., & Caplan, J. B. (2012). Building a memory palace in minutes: Equivalent memory performance using virtual versus conventional environments with the method of loci. *Acta psychologica*, *141*(3), 380–390.

Lewis, J. R., & Sauro, J. (2009). The factor structure of the system usability scale. In *International conference on human centered design*, (pp. 94–103). Springer.

Logie, M. R., & Donaldson, D. I. (2021). Do doorways really matter: Investigating memory benefits of event segmentation in a virtual learning environment. *Cognition*, *209*, 104578.

Lopez, M. C., Deliens, G., & Cleeremans, A. (2016). Ecological assessment of divided attention: what about the current tools and the relevancy of virtual reality. *Revue neurologique*, *172*(4-5), 270–280.

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. *The Cambridge handbook of multimedia learning*, *147*, 158.

Mandler, G. (1967). Organization and memory. In *Psychology of learning and motivation*, vol. 1, (pp. 327–372). Elsevier.

Marschall, S., & Werner, A. (2018). Basics of colour research. *Colour Turn*, (1).

McFadyen, J., Nolan, C., Pinocy, E., Buteri, D., & Baumann, O. (2021). Doorways do not always cause forgetting: a multimodal investigation. *BMC psychology*, *9*(1), 1–13.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321–1329.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Molenberghs, P., Hayward, L., Mattingley, J. B., & Cunnington, R. (2012). Activation patterns during action observation are modulated by context in mirror system areas. *Neuroimage*, *59*(1), 608–615.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of experimental psychology*, *64*(5), 482.

Murf-Studio (2022). go from text to speech with versatile voice generator 2022.
URL `https://murf.ai`

Nolé, M. L., Higuera-Trujillo, J. L., & Llinares, C. (2021). Effects of classroom design on the memory of university students: From a gender perspective. *International Journal of Environmental Research and Public Health*, *18*(17), 9391.

Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical education*, *46*(7), 636–647.

Paes, D., Arantes, E., & Irizarry, J. (2017). Immersive environment for improving the understanding of architectural 3d models: Comparing user spatial perception between immersive and traditional virtual reality systems. *Automation in Construction*, *84*, 292–303.

Persensky, J. J., & Senter, R. (1969). An experimental investigation of a mnemonic system in recall. *The Psychological Record*, *19*(3), 491–499.

Pettijohn, K. A., & Radvansky, G. A. (2016). Walking through doorways causes forgetting: Event structure or updating disruption? *Quarterly Journal of Experimental Psychology*, *69*(11), 2119–2129.

Pettijohn, K. A., & Radvansky, G. A. (2018). Walking through doorways causes forgetting: recall. *Memory*, *26*(10), 1430–1435.

Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., & Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition*, *148*, 136–144.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and instruction*, *12*(1), 61–86.

Quest, M. (2022). Quest 2.
URL `https://www.oculus.com/quest-2/`

Radvansky, G., Andrea, E., & Fisher, J. S. (2017). Event models and the fan effect. *Memory & cognition*, *45*(6), 1028–1044.

Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science*, *21*(4), 269–272.

Radvansky, G. A., & Zacks, J. M. (2011). Event perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 608–620.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current opinion in behavioral sciences*, *17*, 133–140.

Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of experimental psychology: learning, memory, and cognition*, *17*(5), 940.

Ragan, E. D., Sowndararajan, A., Kopper, R., & Bowman, D. A. (2010). The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments*, *19*(6), 527–543.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature reviews neuroscience*, *11*(4), 264–274.

Robin, J., Buchsbaum, B. R., & Moscovitch, M. (2018). The primacy of spatial context in the neural representation of events. *Journal of Neuroscience*, *38*(11), 2755–2765.

Roediger, H. L. (1980). The effectiveness of four mnemonics in ordering recall. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 558.

Rosen, M. A., Salas, E., Pavlas, D., Jensen, R., Fu, D., & Lampton, D. (2010). Demonstration-based training: A review of instructional features. *Human factors*, *52*(5), 596–609.

Ruddle, R. A., Volkova, E., Mohler, B., & Bülthoff, H. H. (2011). The effect of landmark and body-based sensory information on route knowledge. *Memory & cognition*, *39*(4), 686–699.

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, *10*(3), 266–281.

Schuchardt, P., & Bowman, D. A. (2007). The benefits of immersion for spatial understanding of complex underground cave systems. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, (pp. 121–124).

Schwind, V., Knierim, P., Haas, N., & Henze, N. (2019). Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, (pp. 1–12).

Sheffield, F. D. (1961). Theoretical considerations in the learning of complex sequential tasks from demonstration and practice. *Student response in programmed instruction*, *19*(6), 13–32.

Shin, Y. S., Masís-Obando, R., Keshavarzian, N., Dáve, R., & Norman, K. A. (2021). Context-dependent memory effects in two immersive virtual reality environments: On mars and underwater. *Psychonomic Bulletin & Review*, *28*(2), 574–582.

Slater, M. (2014). Grand challenges in virtual environments. *Frontiers in Robotics and AI*, *1*, 3.

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, *3*, 74.

Slater, M., & Sanchez-Vives, M. V. (2022). Is consciousness first in virtual reality? *Frontiers in Psychology*, *13*.

Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition*, *10*(5), 405–412.

Smith, S. M. (1984). A comparison of two techniques for reducing context-dependent forgetting. *Memory & Cognition*, *12*(5), 477–482.

Smith, S. M. (1985). Effects of number of study environments and learning instructions on free-recall clustering and accuracy. *Bulletin of the Psychonomic Society*, *23*(6), 440–442.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342–353.

Smith, S. M., & Manzano, I. (2010). Video context-dependent recall. *Behavior Research Methods*, *42*(1), 292–301.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, *8*(2), 203–220.

Stangl, M., Achtzehn, J., Huber, K., Dietrich, C., Tempelmann, C., & Wolbers, T. (2018). Compromised grid-cell-like representations in old age as a key mechanism to explain age-related navigational deficits. *Current Biology*, *28*(7), 1108–1115.

SteamVR (2017). Steamvr - introducing steamvr home beta - steam news. URL `https://store.steampowered.com/news/app/250820/view/2898585530113881631`

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257–285.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1589.

Urgolites, Z. J., & Wood, J. N. (2013). Visual long-term memory stores high-fidelity representations of observed actions. *Psychological science*, *24*(4), 403–411.

Van Merrienboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review*, *17*(2), 147–177.

Vive (2022). Find the right high-end vr system for you. URL `https://www.vive.com/uk/product/#vive-spec`

Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron*, *35*(1), 205–211.

Wälti, M. J., Woolley, D. G., & Wenderoth, N. (2019). Reinstating verbal memories with virtual contexts: Myth or reality? *PloS one*, *14*(3), e0214540.

Ward, G. (2002). A recency-based account of the list length effect in free recall. *Memory & Cognition*, *30*(6), 885–892.

Watson, P. (2023). Supplementary material. URL `https://pearl.plymouth.ac.uk/bitstream/item/ec69f307-cf48-4c8f-8a1e-a4ec7f693b5f/2023watson10554952phd_SupplementaryMaterial.zip`

Watson, P., & Gaudl, S. E. (2021). Walking through virtual doors: A study on the effects of virtual location changes on memory. *International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments*.

Watson, P., & Livingstone, D. (2018). Using mixed reality displays for observational learning of motor skills: A design research approach enhancing memory recall and usability. *Research in Learning Technology*, *26*.

Westfall, J. (2015). Pangea: Power analysis for general anova designs. *Unpublished manuscript. Available at http://jakewestfall. org/publications/pangea. pdf*.

Williams, A. M., & Hodges, N. J. (2005). Practice, instruction and skill acquisition in soccer: Challenging tradition. *Journal of sports sciences*, *23*(6), 637–650.

Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological review*, *105*(3), 558.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, *7*(3), 225–240.

Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: an instance of the ideomotor principle. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 501–515.

Wulf, G., & Schmidt, R. A. (1997). Variability of practice and implicit motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 987.

Yang, Y., Leung, H., Yue, L., & Deng, L. (2013). Generating a two-phase lesson for guiding beginners to learn basic dance movements. *Computers & Education*, *61*, 1–20.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological bulletin*, *127*(1), 3.

Zanbaka, C. A., Lok, B. C., Babu, S. V., Ulinski, A. C., & Hodges, L. F. (2005). Comparison of path visualizations and cognitive measures relative to travel technique in a virtual environment. *IEEE Transactions on Visualization and Computer Graphics*, *11*(6), 694–705.