CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	The colour of sound: enhancing pitch-colour cross-modal associations in virtual reality
Authors	O'Toole, Patrick
Publication date	2025-07-31
Original Citation	O'Toole, P. 2025. The colour of sound: enhancing pitch-colour cross-modal associations in virtual reality. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2025, Patrick O'Toole. - https://creativecommons.org/licenses/by-nc/4.0/
Download date	2026-10-26 08:15:49
Item downloaded from	https://hdl.handle.net/10468/18824



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ollscoil na hÉireann, Corcaigh
National University of Ireland, Cork



The Colour of Sound: Enhancing Pitch-Colour Cross-Modal Associations in Virtual Reality

Thesis presented by

Patrick O'Toole, B.A.

ORCID: 0000-0001-6130-1024

for the degree of Doctor of Philosophy

University College Cork
School of Computer Science & Information Technology
Head of School/Department: Prof Dirk Pesch
Supervisor(s): Dr Laura Maye, Dr Ian Pitt & Dr Maurizio Mancini

2025

Abstract

This thesis investigates how Virtual Reality (VR) can enhance cross-modal associations between pitch and colour. While previous research has shown that people often associate higher pitches with brighter colours and lower pitches with darker ones, it remains unclear whether immersive VR environments can provide a more meaningful experience that could support, personalise, or strengthen these associations. The potential of VR is how key immersive features, such as multisensory integration, embodied interaction, spatial audio, and affective environments, may influence the formation of pitch-colour associations in a more engaging and meaningful way.

To address this, three mixed-methods studies were conducted, each progressing the understanding of pitch-colour cross-modality in VR. Study One tested whether cross-modal consistency scores differ between desktop and VR conditions. Study Two explored the role of immersive VR features, such as environment design, interaction, and spatial audio, in shaping users' engagement and experience of pitch-colour associations. Study Three evaluated whether personalised pitch-colour cross-modal mappings, created by participants, offered any potential learning benefits over predefined mappings in a music theory task.

Findings from across the studies indicate that VR is a viable platform for investigating pitch-colour associations, and that design elements such as multisensory integration, interactions, embodiment, and affective environment cues can support user engagement and perceptual clarity. While consistency scores did not differ significantly between environments in Study One, participants expressed a strong preference for VR, citing greater immersion and focus. Studies Two and Three highlighted the use of minimalistic and affective environments aided task engagement, and that participants benefited from interactive tasks that supported embodied cognition.

Taken together, the studies suggest that immersive VR environments, when designed with attention to cross-modal coherence, interactivity, and affect, can support a deeper understanding of pitch-colour associations. This has potential for application in learning contexts, such as the music theory training explored in Study Three. The findings have implications for the design of future VR tools in music education, multisensory learning, and accessible interface design, and also point towards further research into adaptive personalisation in virtual learning environments.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Laura Maye, Ian Pitt and Maurizio Mancini, for their invaluable guidance, support, and encouragement throughout this journey. I am also thankful to Robert Bresin, Kjetil Falkenberg, David Murphy, Dr Donald Glowinski, Dr Jason Chan, and Andrea Chirico for their insightful feedback and constructive advice, which have greatly enriched this work.

I would like to acknowledge the contributions of ADVANCE CRT program, especially Luis Gómez de Membrillera, James Duggan, and Dirk Pesch, whose assistance was crucial in completing this research.

Special thanks to my fellow PhD colleagues and friends, for their camaraderie and support. Finally, my heartfelt appreciation goes to my family for their unwavering love and belief in me.

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism and intellectual property.

Signed: Patrick O'Toole

Table of Contents

Abstract i
Acknowledgements ii
Declaration iii

List of Tables **ix**

List of Figures **x**

1 Introduction **1**
1.1 Motivation 3
1.2 Background 4
1.2.1 Cross-Modal Associations and Synaesthesia 4
1.2.2 Immersive Virtual Reality Environments 5
1.2.3 In the Context of Music Learning 6
1.3 Research Questions 7
1.4 Contribution to the Field 9
1.5 Overview of the Thesis 10

2 Literature Review **12**
2.1 Literature Review Methodology 12
2.2 Cross-Modal Associations 13
2.2.1 What are Cross-Modal Associations? 13
2.2.2 Auditory-Visual Cross-Modal Associations 15
2.2.3 Relationship with Synaesthesia 17
2.3 Synaesthesia 18
2.3.1 Differences in Synaesthesia 18
2.3.2 Chromaesthesia or Sound-Colour Synaesthesia 19
2.3.3 Testing for Synaesthesia 20
2.4 VR Technologies and Capabilities 24
2.4.1 Immersion 24
2.4.2 Presence 26
2.4.3 Interactions in VR 29
2.4.4 Spatial Audio 30
2.5 Multisensory Learning and Engagement 32
2.5.1 Multisensory Integration and Cross-Modality 33
2.5.2 Embodied Cognition in VR 34
2.5.3 Affective Factors in Multisensory Learning 35
2.5.4 Cognitive Load Theory (CLT) 35
2.5.5 Cognitive Theory of Multimedia Learning (CTML) 36
2.5.6 VR as a Tool for Cross-Modal Learning 36

2.6 Cross-Modal Research in XR Environments . . . 38
2.6.1 VR as a Platform for Multisensory & Cross-Modal Experience 38
2.6.2 Evidence of Cross-Modal Correspondences in XR Settings . . 39
2.6.3 Challenges & Considerations for Cross-Modal VR . . . 40
2.7 Conclusion . . . 41

3 Methodology 43
3.1 Introduction . . . 43
3.2 Epistemological Positioning . . . 43
3.3 Research Design and Rationale . . . 44
3.3.1 Study One: . . . 44
3.3.2 Study Two: . . . 45
3.3.3 Study Three: . . . 45
3.4 Mixed Methods Approach . . . 45
3.4.1 Quantitative Methodologies . . . 46
3.4.2 Qualitative Methodology . . . 47
3.5 Overview of Methods Used in Each Study . . . 48
3.6 Ethical Considerations . . . 50
3.7 Chapter Reflection . . . 50

4 Pitch-Colour Associations in VR vs Desktop 51
4.1 Introduction . . . 52
4.2 Related Work . . . 52
4.2.1 Auditory-Visual Cross-Modal Associations . . . 53
4.2.2 Current Tests on Auditory-Visual Associations . . . 53
4.2.3 Similar Virtual Reality Research . . . 54
4.3 Methodology . . . 54
4.3.1 Experimental Design . . . 55
4.3.2 Participants . . . 55
4.3.3 Data Collection . . . 56
4.3.4 Data Analysis . . . 56
4.4 Results . . . 57
4.5 Discussion . . . 57
4.5.1 Implications & Limitations . . . 58
4.6 Conclusion . . . 58
4.7 Delving Deeper into Auditory-Visual Tests in VR & Desktop Environments . . . 59
4.8 Additional Data Analysis . . . 59
4.8.1 Colour brightness & distance choices . . . 59
4.8.2 Qualitative Analysis . . . 60
4.9 Additional Results . . . 60
4.9.1 Colour brightness & distance choices . . . 60
4.9.2 Post-Study Questionnaire Analysis . . . 61
4.10 Additional Discussion . . . 67
4.10.1 Effect of Environment (Desktop vs VR HMD) . . . 67
4.10.2 Insights for Future Improvements . . . 69
4.10.3 Additional Observations: . . . 71
4.11 Final Conclusion . . . 72
4.12 Chapter Reflection . . . 72

5 Immersive VR Pitch-Colour Study **74**
5.1 Introduction . . . 75
5.2 Related Work . . . 77
5.2.1 Auditory-Visual Associations . . . 77
5.2.2 Multi-Modal Experiences in Virtual and Extended Reality . . . 78
5.2.3 Immersion, Presence, & Interaction in VR . . . 79
5.2.4 Virtual Environments & User Experience . . . 80
5.3 Methodology . . . 81
5.3.1 Environment Setup & Design . . . 82
5.3.2 Captured Data . . . 84
5.3.3 Experimental Design . . . 85
5.3.4 Participants . . . 86
5.3.5 Data Collection & Analysis . . . 87
5.4 Results . . . 87
5.4.1 General Overview . . . 88
5.4.2 Standardised Assessment Scales . . . 88
5.4.3 Completion Times & Colour-Pitch/Size Association Analysis . . . 90
5.4.4 Template Analysis Results . . . 90
5.5 Discussion . . . 96
5.5.1 RQ1: How do the two VR environments impact user experience and performance in the pitch-colour association task? . . . 96
5.5.2 RQ2: How do changes between the two environments affect participants' experiences? . . . 98
5.5.3 RQ3: What improvements would make the system more effective for the pitch-colour association task? . . . 99
5.6 Limitations & Future Work . . . 101
5.7 Conclusion . . . 102
5.8 Chapter Reflection . . . 103

6 Cross-Modal Music Learning in VR **105**
6.1 Introduction . . . 106
6.2 Background . . . 108
6.2.1 Pitch-Colour Cross-Modal Associations . . . 108
6.2.2 VR as a Learning Tool: Designing Immersive and Cross-Modal Environments . . . 109
6.2.3 Musical Education . . . 110
6.2.4 Current VR Music Applications . . . 111
6.3 Methodology . . . 114
6.3.1 Participants . . . 114
6.3.2 Apparatus & Stimuli . . . 114
6.3.3 Experimental Design . . . 115
6.3.4 Data Collection . . . 117
6.3.5 Data Analysis . . . 117
6.4 Results . . . 119
6.4.1 Quantitative Analysis Results . . . 119
6.4.2 Qualitative Analysis Results . . . 123
6.5 Discussion . . . 134
6.5.1 Research Question One . . . 134

6.5.2 Research Question Two . . . 135
6.5.3 Research Question Three . . . 137
6.5.4 Research Question Four . . . 137
6.6 Limitations & Future Work . . . 139
6.7 Conclusion . . . 140
6.8 Chapter Reflection . . . 141

7 Conclusions and Future Work 143
7.1 Summary of Findings . . . 144
7.1.1 Study One . . . 145
7.1.2 Study Two . . . 145
7.1.3 Study Three . . . 146
7.2 Themes Across Studies . . . 147
7.2.1 Iterative Design Evolution . . . 147
7.2.2 Impact of VR Features: Environment, Interaction and Immersion on Pitch-Colour Tasks . . . 148
7.2.3 Personalisation vs Pre-Defined Cross-Modal Associations . . . 149
7.2.4 Cross-Modal Integration . . . 150
7.2.5 Impact of Affective Environments on Pitch-Colour Associations 151
7.2.6 Individual Differences in Cross-Modal Learning . . . 153
7.3 Contributions . . . 154
7.3.1 Practical Contributions . . . 154
7.3.2 Theoretical Contributions . . . 157
7.4 Can VR Enhance Cross-Modality? . . . 159
7.5 Limitations . . . 160
7.6 Future Work . . . 162
7.7 Conclusion . . . 162

Bibliography 164

A Study One 191
A.1 Consent Form . . . 191
A.2 Information Sheet . . . 192
A.3 Study One Ethics Approval . . . 193
A.4 Pre-Study Questionnaire Items . . . 194
A.5 Post-Study Questionnaire Items . . . 195
A.6 Evolution of Template Analysis Themes - Study One . . . 196

B Study Two 197
B.1 Consent Form . . . 197
B.2 Information Sheet . . . 198
B.3 Study Two Ethics Approval . . . 200
B.4 Pre-Study Questionnaire Items . . . 201
B.5 Post-Study Questionnaire Items . . . 202
B.6 System Usabilty Scale (SUS) . . . 203
B.7 iGroup Presence Questionnaire (IPQ) . . . 204
B.8 Simulator Sickness Questionnaire (SSQ) . . . 205
B.9 Evolution of Template Analysis Themes - Study Two . . . 206
B.10 Images of the VR Application in Study Two . . . 207

C Study Three **209**
C.1 Consent Form 209
C.2 Information Sheet 210
C.3 Study Three Ethics Approval 212
C.4 Pre-Study Questionnaire Items 213
C.5 Post-Study Questionnaire Items 215
C.6 Evolution of Template Analysis Themes - Study Three 216
C.7 Images of the VR Application in Study Three 217
C.8 SPSS Output from Study Three 219

D Miscellaneous **221**
D.1 Colour Comparison Between a VR HMD and Desktop 221
D.2 Test of an Alternative Browser in VR 222
D.3 VR HMD Specs for the Quest 2 and Newer 223

E Publications and Contributions **224**

List of Tables

2.1 Cross-Modal Correspondences 16

3.1 Study Designs and Methods Across the Three Studies 49

5.1 Study Two: Wilcoxon Test Results and Mean Scores 88
5.2 Study Two: Simulator Sickness Questionnaire (SSQ) Results 88

6.1 Study Three: Participant Demographics and Group Split 113
6.2 Study Three: Results on Error Scores and Time Taken 120
6.3 Significant Bonferroni-corrected pairwise comparisons for error scores. L = Lesson in the Comparison column. 95% CI shown as [lower, upper]. "Direction" indicates which lesson had higher errors. 121
6.4 Significant Bonferroni-corrected pairwise comparisons for completion time. L = Lesson in the Comparison column. 95% CI shown as [lower, upper]. "Direction" indicates which lesson took longer. 121

D.1 Comparison of VR Headsets Released Between 2020-2024 223

List of Figures

1.1 Kandinsky's Composition VI 2
1.2 Central Thesis Research Question and Sub-Questions 8

2.1 Synaesthesia as a Continuum 17

4.1 Study One: Testing Environment 55
4.2 Study One: Study Design 56
4.3 Study One: Pitch-Brightness Associations for VR and Desktop Settings 61
4.4 Study One: Colour-Notes Distances 62
4.5 Study One: Environment Preferences and Synaesthesia Responses Pie Charts 62
4.6 Study One: Test Difficulty and VR Experience 63
4.7 Study One: Participant Feedback Text Links 63
4.8 Study One: Participant Feedback Word Cloud 64

5.1 Study Two: Test Environment 81
5.2 Study Two: The Four VR Scenes 82
5.3 Study Two: Study Design 86
5.4 Study Two: Completion Times per Environment 89
5.5 Study Two: Average Times per Pitch-Colour Interaction 90
5.6 Study Two: Pitch-Colour Frequencies Heat Map 91
5.7 Study Two: Pitch-Size Frequencies Heat Map 91

6.1 Study Three: Scriabin's Circle of Fifths 111
6.2 Study Three: Test Environment 114
6.3 Study Three: Study Design 116
6.4 Study Three: The Eight Music Lessons 118
6.5 Study Three: Mean Error Scores Across Lessons by Group 122
6.6 Study Three: Mean Time Taken Across Lessons by Group 122
6.7 Study Three: Heat Map of Colour Choices by Note (C4-C5) 123
6.8 Heat Map of Colour Family Selections by Note 124
6.9 Study Three: Frequency of Colour Brightness Choices per Note . . . 125

A.1 Ethics Approval Email 193
A.2 Study One: Template Analysis Coding 196

B.1 Ethics Approval Email 200
B.2 Study Two: Template Analysis Coding 206
B.3 Study Two: Early Iteration of Forest Visual 207
B.4 Study Two: Early Iteration of Space Visual 207

B.5 Study Two: First Iteration Forest Tutorial 208
B.6 Study Two: First Iteration Space Tutorial 208

C.1 Ethics Approval Email 212
C.2 Study Three: Template Analysis Coding 216
C.3 Study Three: Early iteration of the Music Exercises 217
C.4 Study Three: Pre-defined Pitch-Colour Mappings 217
C.5 Study Three: Final Scores Menu 218
C.6 Study Three: Visual Feedback 218
C.7 Within-subjects effect on Error Scores 219
C.8 Within-subjects Contrast on Error Scores 219
C.9 Between-subjects Effect on Error Scores 220
C.10 Within-subjects effect on Time Taken 220
C.11 Within-subjects Contrast on Time Taken 220
C.12 Between-subjects Effect on Time Taken 220

D.1 Colour Comparison Between a VR HMD and Desktop 221
D.2 Test of an Alternative Browser in VR 222

Acronyms

2D Two Dimensional. 6

3D Three Dimensional. 6

ADHD Attention-Deficit/Hyperactivity Disorder. 162

AI Artificial Intelligence. 10, 140

AIFF Audio Interchange File Format. 115

AR Augmented Reality. 24

CAVE Cave Automatic Virtual Environment. 24

CLM Cognitive Load Measurement. 46

CLT Cognitive Load Theory. 33

CSV Comma-Separated Values. 56

CTML Cognitive Theory of Multimedia Learning. 33

FOR Fields of Regard. 24

FOV Fields of View. 24

HCI Human-Computer Interaction. 3

HMD Head-Mounted Display. 1

HRTF Head-Related Transfer Functions. 31

HSV Hue, Saturation, and Value. 60

IAT Implicit Association Test. 21

IPA Interpretative Phenomenological Analysis. 48

IPQ iGroup Presence Questionnaire. 27

ITU-R International Telecommunication Union Radiocommunication. 59

JSON JavaScript Object Notation. 87

K-12 Kindergarten to twelfth Grade. 136

K-6 Kindergarten to Sixth Grade. 109

LCD Liquid Crystal Display. 55

MIDI Musical Instrument Digital Interface. 115

MP3 MPEG-1 Audio Layer 3. 115

MR Mixed Reality. 24

NeCoSyn Netherlands Colour Synaesthesia. 21

RGB Red, Green, Blue Colour Model. 46

RR Reduced Realities. 106

RTS Reaction Times. 15

SDK Software Development Kit. 115

SSQ Simulator Sickness Questionnaire. 46

SUS System Usability Scale. 46

SUS Slater-Usoh-Steed Questionnaire. 27

TOG Test of Genuineness. 21

TPO Temporo-Parietal-Occipital. 18

UI User Interface. 80

URP Universal Render Pipeline. 115

VE Virtual Environment. 1

VR Virtual Reality. 1

WS Witmer and Singer Questionnaire. 27

XR Extended Reality. 12

XRI XR Interaction Toolkit. 161

Chapter 1

Introduction

Wassily Kandinsky, a pioneer of abstract art and a theorist on the connections between sensory modalities, once wrote:

> *"Colour is the keyboard, the eyes are the hammers, the soul is the piano with many strings. The artist is the hand which plays, touching one key or another, to cause vibrations in the soul."* (Kandinsky, 2022)

This passage poetically illustrates the interplay between visual and auditory perception, drawing a parallel between colour, music, and emotion. Kandinsky likens colour to a piano key, waiting to be struck, with the eye as the hammer that perceives it and the artist's hand as the catalyst that brings it to life. Just as a sequence of notes can evoke emotions, the interaction of colour and sound can shape perception and emotional experience.

Kandinsky's perspective resonates with the central focus of this thesis: exploring the associations between musical pitch (a perceptual property related to the frequency of sound waves (Oxenham, 2012)), and colour (a visual perception resulting from the interpretation of light wavelengths by the visual system (Kalloniatis and Luu, 1995)). While pitch-colour associations have been explored in various contexts—most notably through pen-and-paper consistency tests (Baron-Cohen et al., 1987), desktop-based colour selection tools (Asher et al., 2006), and online platforms such as the `Synaesthesia Battery` (Eagleman et al., 2007)—these approaches are typically limited to two-dimensional, non-interactive environments. The connection between sound and visual representation dates back centuries, with early *"colour-music"* systems emerging as early as the 1500s. One of the most well-known examples is Alexander Scriabin's 1910 composition *Prometheus: The Poem of Fire*, which featured a *"colour organ"* that projected synchronised coloured lights as part of the performance (Rucsanda, 2017).

While Scriabin used the technologies of his time to create a multi-sensory experience, today's advances offer new possibilities. In this thesis, the term Virtual Environment (VE) describes interactive, computer-generated spaces. This term can refer to VEs more generally, or to ones specifically built and experienced in Virtual Reality (VR). VR refers to immersive systems that let users explore and interact with these environments, usually through a Head-Mounted Display (HMD). A VE provides a means to explore pitch-colour associations in controlled, immersive, and interactive contexts. This thesis by publication investigates how features such as immersion, presence, interaction, and spatial audio in VEs can deepen our understanding of auditory-visual relationships, specifically between pitch and colour, and



Figure 1.1: Composition VI by Wassily Kandinsky, 1913. Image sourced from Timeless Moon Archive. Public domain.

enhance how these multi-modal associations are experienced. VR is an inherently multi-modal technology, with extensive research conducted in music development (Trujano et al., 2018; Molero et al., 2021), social and natural sciences (Pellas et al., 2020), and accessibility education (Lorenzo et al., 2022; Rudenko and Haahr, 2022), to name a few.

The terms ‘‘enhance’’, ‘‘enhancement’’, and ‘‘enhancing’’ are used throughout this thesis in different contexts. It is used to describe the ways in which immersive VR features can potentially support, shape, or strengthen cross-modal associations and experiential qualities of the interactions. The research further considers how cross-modal design may influence users' performance and understanding of specific concepts, particularly in relation to exploratory learning contexts. Studies of enhancement in VR and HCI emphasise the role of design, informed by usability, emotion, and user experience, and highlight how aligning interaction design with learning theory can support richer forms of understanding (Ahn and Clegg, 2018; Zhou et al., 2018). Chapters 4, 5, and 6 present a series of design-led investigations into the implementation of pitch-colour cross-modal associations in VR. Perceptual accuracy and users' understanding of musical concepts are also examined, but primarily as a means of evaluating the impact of the design interventions, rather than as claims of perceptual or educational improvement. The thesis therefore adopts an experiential, design-oriented perspective, examining how design choices in VR can shape the formation and expression of pitch-colour cross-modal relationships.

1.1 Motivation

My motivation for the research presented in this thesis stems from my passion for playing music and exploring new forms of creative expression. A visit to the Hermitage Museum in St. Petersburg, Russia, was pivotal in shaping my interest in the relationship between music and colour. While there, I encountered Kandinsky's `Composition IV` (see Figure 1.1), a work that profoundly shifted my perspective on how our senses communicate with each other to form a perception of a piece of art. I acquired a print of `Composition VI` that day. Whenever I reflect on it, I sense a profound expression of musical interplay, with various instruments, melodies, and rhythms emerging through its intricate interplay of colours, shapes, and textures.

As Kandinsky was known to have synaesthesia, this perceptual phenomenon intrigued me, prompting an exploration of synaesthesia and the broader field of cross-modal associations. Synaesthesia is a phenomenon in which stimulation of one sensory modality triggers an involuntary response in another (Cytowic, 2003). In contrast, cross-modal associations occur when one sensory modality automatically evokes or reinforces another, but without the involuntary nature of synaesthesia (Spence, 2011). The key differences between the two lie in the degree of automaticity, consistency, and specificity of the associations (Ward, 2013; Curwen, 2018).

This initial interest in synaesthesia led me to explore the broader field of cross-modal associations, particularly the relationship between pitch and colour. While synaesthetes (a person with synaesthesia) naturally experience these connections, research suggests that non-synaesthetes (a person who does not experience synaesthesia) may also exhibit systematic but weaker pitch-colour associations (Ward, 2013). From a Human-Computer Interaction (HCI) perspective, the relationship between synaesthesia and cross-modality can provide an interesting avenue of exploration. Understanding multi- and cross-modal connections in VR can provide insights into how different sensory modalities interact, shaping how interactive systems facilitate learning and engagement.

Researchers in the field of HCI have extensively studied multi-modality to enhance user experience by integrating multiple sensory channels, such as visual, auditory, and haptic inputs (Marcus and Rosenzweig, 2020). In the context of HCI, multi-modality involves presenting information through multiple sensory channels (e.g., sight and sound) (Jacko and Sears, 2003). However, while multi-modality—the presentation of multiple sensory modalities in parallel, such as combining audio narration with on-screen text—has been widely explored, cross-modality—the interaction between senses, such as associating a musical pitch with a colour—has received comparatively less attention, particularly in the context of VR (Malpica et al., 2020). This topic will be further explored in Section 2.2.

Cross-modal research can play a significant role in HCI by informing the design of interfaces that leverage sensory integration for improved user experience. Its applications extend to accessibility and assistive technologies, such as software that visualises music through colour for users with hearing impairments (Szucs et al., 2018); affective computing, such as cross-modal physiological input for improved emotion recognition (Bhatti et al., 2021); and immersive interaction design, such as systems used to train neural plasticity in ageing brains (Rudenko and Cabral, 2019).

Studies have demonstrated that cross-modal cues can enhance perception, engagement, and learning within interactive systems (Deroy and Spence, 2016), yet

much of this research has been conducted in non-immersive environments. Given the increasing adoption of VR, there is a growing need to explore how people can utilise immersive technologies to strengthen cross-modal associations, particularly concerning auditory-visual correspondences such as pitch-colour mapping.

The benefits of cross-modal design align closely with VR's ability to offer dynamic, multi-sensory experiences. By providing real-time interaction with auditory and visual stimuli, VR presents a unique opportunity to examine how carefully designed immersion can reinforce perceptual associations. This thesis builds on these principles to investigate whether thoughtfully designed VR can enhance cross-modal pitch-colour correspondences, bridging the gap between natural synaesthetic perception and structured cross-modal learning.

Understanding these associations in HCI has practical implications for music education, accessibility, and interactive system design. Insights from this research could contribute to developing sensory-rich learning tools, assistive technologies, and adaptive interfaces that enhance user experience across various domains.

1.2 Background

The three main research areas that form the foundation of this thesis are cross-modal pitch-colour associations, immersive virtual reality environments, and the potential applications for VR technology in music learning. These areas are each reflected across the three studies presented in Chapters 4, 5, and 6. Each area contributes to understanding how multi-sensory experiences influence perception, cognition, and learning.

1.2.1 Cross-Modal Associations and Synaesthesia

Cross-modal associations (explained in Section 1.1) exist across various sensory modalities, including links between auditory and visual features (Spence, 2011). In the case of pitch and colour, research has shown that higher pitches are often associated with brighter or lighter colours, while lower pitches correspond to darker shades (Marks, 1987). Although these associations are widely observed in the general population, they differ from synaesthesia, a condition in which such connections are involuntary and remain consistent throughout a person's lifetime (Cytowic, 2003; Ward, 2013). Due to these similarities, cross-modal correspondences are sometimes referred to as *"weak synaesthesia"* (Martino and Marks, 2001), as they exhibit comparable sensory patterns but lack the automaticity and high consistency of synaesthetic experiences.

To assess pitch-colour associations, researchers have developed various consistency tests and methods for studying associations across other sensory modalities (Eagleman et al., 2007; Asher et al., 2006; Campen and Froger, 2003). Traditionally, these studies have used pen-and-paper or desktop-based tasks in which participants match colours to auditory pitches, with results used to identify synaesthetic consistency patterns. While these tests effectively diagnose sound-colour synaesthesia and examine differences between synaesthetes and non-synaesthetes, research into pitch-colour associations has potential applications beyond diagnostic purposes.

Research into synaesthesia and sound-colour cross-modal associations has produced valuable insights across various fields, including art and creativity, education,

and assistive technologies. From a creative perspective, exploring sound-colour correspondences has led to new forms of multi-modal artistic expression. Studies have examined synaesthesia as a potential gateway to creativity or as a means of cultivating it (Murgia, 2015; Merter, 2017), with Haverkamp (2022) specifically investigating synaesthetic art.

In education, pitch-colour associations have been explored as a way to reinforce learning by engaging multiple senses. For example, multi-sensory augmented reality (AR) has been used to teach students about Kandinsky's artwork (Belloli et al., 2020), and a sound-colour notation system has been proposed for beginner musicians (Kuo and Chuang, 2013). Another area benefiting from research into sound-colour associations is assistive technology. A notable example is `The vOICe`, a sensory substitution device that converts visual images into auditory signals, potentially constituting a form of acquired synaesthesia (Ward and Meijer, 2010).

Prior to commencing this PhD, I developed an application called `SoundStrokes` to investigate whether participants could strengthen pitch-colour cross-modal associations through practice. A study using `SoundStrokes` found that participants' pitch-colour associations became more consistent over time, suggesting that exposure and interaction can reinforce these connections (O'Toole et al., 2019). The study also highlighted the role of emotion in shaping auditory-visual associations. A follow-up study by O'Toole et al. (2021) (P1) further examined sound-emotion associations, investigating how musical experience, personality traits, and learning styles influenced participants' choices. The results confirmed a fundamental emotional link to pitch, with lower notes commonly associated with *"sadness"* and higher notes with *"joy"*, aligning with the initial findings of O'Toole et al. (2019). This initial work emphasised the complexity of pitch-colour associations, suggesting that factors such as engagement, emotion, musical experience, personality, and learning styles can contribute to understanding how cross-modal correspondences can be used effectively in music learning contexts. The range of factors highlighted above suggests that designing cross-modal learning tools should account for individual differences, potentially allowing for personalised or adaptive systems that respond to user preferences or learning profiles.

Despite these advancements, traditional synaesthesia consistency tests and cross-modal association studies often lack immersive and interactive elements, which could strengthen the connection between sensory modalities and improve learning effectiveness. VR technologies present an opportunity to address these limitations by offering dynamic, interactive environments that could be designed to enhance the formation and retention of pitch-colour associations. The following section explores the role of immersive VR in cross-modal research and its potential applications in learning contexts.

1.2.2 Immersive Virtual Reality Environments

Using immersive technology, such as VR, presents a novel approach to exploring pitch-colour cross-modal associations in a dynamic, spatially engaging setting. VR allows for real-time manipulation of auditory and visual stimuli, providing a richer, multi-sensory experience that could strengthen cross-modal understandings.

Where the focus of VR has mainly been directed at multi-modal experiences, less research has been conducted on cross-modality and its benefits on providing

immersive and engaging experiences (Malpica et al., 2020). Studies have demonstrated that immersive settings, such as VR, can improve performance on specific cognitive and perceptual tasks, increase user engagement, and facilitate learning in interactive settings (McMahan et al., 2006; Dede, 2009).

With studies by O'Toole et al. (2021, 2019) highlighting the role of emotion on pitch and colour cross-modal associations, affect on a broader level might be worth exploring. VEs could be designed to foster an initial sense of ease and emotional comfort in a virtual space. A study by Riva et al. (2007) highlights how VR can be viewed as an affective medium, with presence—commonly defined as the psychological sense of *"being there"* in a virtual environment—increasing in *"emotional"* environments. Strategic affective design could play a role in creating the optimal environment for forming and enhancing pitch-colour associations.

In addition to environmental design, other features of VR can contribute to providing an engaging and immersive experience, such as interactions, locomotion, and auditory presentation. With interactions and locomotion, a trade-off usually occurs between comfort, physical space, and virtual vs physical objects to provide an engaging experience when space and cost are factors (Mine et al., 1997). With VRs' capability to present Three Dimensional (3D) audio, several studies have found that this feature has provided more immersive experiences to users compared to Two Dimensional (2D) audio presentation (Broderick et al., 2018).

1.2.3 In the Context of Music Learning

This thesis investigates whether immersive VR environments can enhance the formation and understanding of pitch-colour cross-modal associations compared to traditional methods. One of the main use cases for congruent pitch-colour cross-modal associations is in music education.

Several studies have explored research on cross-modal associations or synaesthesia in music performance and education (Vines et al., 2011; Yau and Mccrindle, 2007), and other studies have explored VR as an effective medium for music learning (Hirci et al., 2023; Timoney et al., 2018). However, few studies have combined these areas to use immersive VR systems to enhance pitch-colour cross-modal associations to provide an effective learning platform for music theory. Kuo and Chuang (2013) proposed a colour notation system for beginner musicians to support music reading without learning complex traditional notation. While Lewis (2020) argues that the current generation of children (*"Gen Alpha"*), shaped by increasingly digital learning environments, may be better suited to multi- or cross-modal approaches to music education

In terms of VR research that has explored music teaching, studies have explored multi-modal VR environments for learning purposes; however, they did not use cross-modal pitch-colour associations. Timoney et al. (2018) proposed an application to use interactive *"singing blocks"* to learn harmony and chords. No studies have directly built upon the proposal of the *"singing blocks"* application; however, other studies have shown promise. For example, `ChordAR`, an application used to teach children music chords, demonstrated that learners could master playing the game and accurately imitate the C, G, and E chords (Lu et al., 2022).

This section has outlined the primary research areas underpinning this thesis. It introduced the concepts of multi-modality, cross-modal associations, and synaes-

thesia, highlighting their relevance for understanding perception and their potential in supporting learning, creativity, and assistive technologies. It then touched on the role of immersive virtual environments (VEs), noting how features like sensory engagement, interaction, and affective design make VR a promising space for exploring auditory-visual relationships such as pitch-colour associations. Lastly, the use of cross-modal design in educational contexts was explored, with music learning presented as one example. Together, these areas point to a gap in how immersive VR experiences can be designed to support pitch-colour associations. The following section presents the research questions that address this gap by investigating the role of VR in facilitating pitch-colour cross-modal associations, with music theory learning serving as a use case.

1.3 Research Questions

This is a thesis by publication. Chapter 4 extends a paper published at `The International Conference on Advanced Visual Interfaces (AVI)` (see P3 in Appendix E) by adding additional quantitative and qualitative results, as well as a more in-depth discussion. Chapter 5 is presents a paper published in the `Virtual Reality Springer Journal` (see P7 in Appendix E), and Chapter 6 has been submitted to the *Computers & Education: X Reality Journal* (see P8 in Appendix E).

The overarching research question guiding this thesis is:

How can the cross-modal associations of pitch and colour be enhanced through virtual reality?

To address this question, the research is structured around three key studies, each investigating different aspects of pitch-colour associations, immersive virtual environments, and their application in music learning. These studies form the core chapters of the thesis. Figure 1.2 presents how each study contributes to answering the overall research question.

The specific research questions for each study are:

- **Chapter 4 (Study 1) - Sound and Colour: Evaluating Auditory-Visual Tests in Virtual Reality and Traditional Desktop Settings**
  - RQ1: Is VR a viable platform for conducting pitch-colour association tests compared to desktop environments?

  This first study (P5) explores a non-immersive VR set-up compared to a standard desktop set-up when taking a pitch-colour consistency test. A firmer understanding of pitch-colour associations in a low-immersion VR setting and how participants form these connections opens up an investigation into which immersive features could further enhance the design the VR environments to effectively present cross-modal connections.

- **Chapter 5 (Study 2) - Learning Musical Pitch-Colour Associations in Immersive Virtual Reality Environments**
  - RQ2: What role do immersive features play in supporting pitch-colour cross-modal associations in VR learning environments?

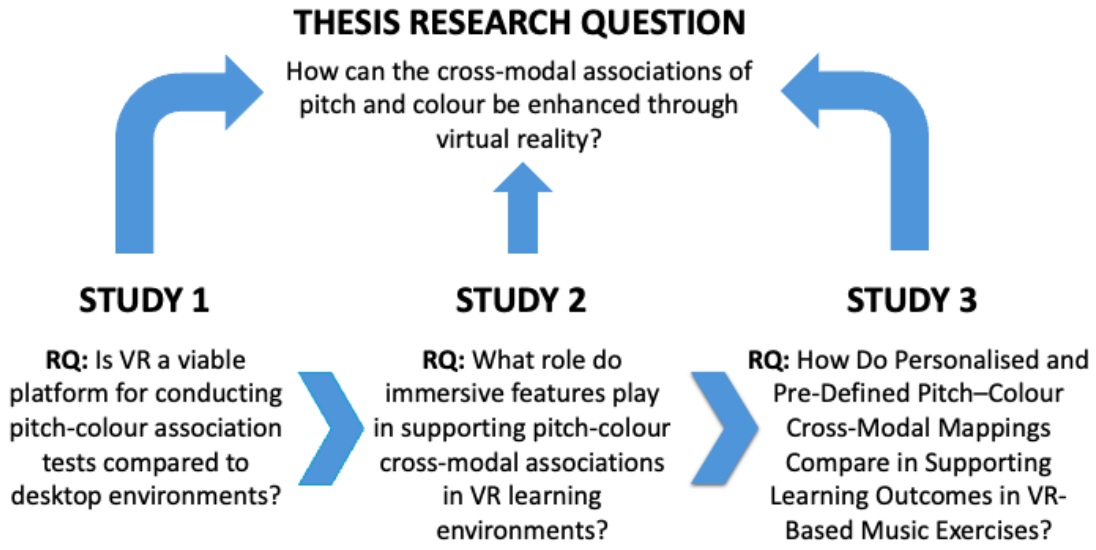


Figure 1.2: This figure presents the central research question at the top, with each study's sub-questions shown below, illustrating how the individual studies address the overall research aim.

After examining how participants formed pitch-colour associations in a low-immersion VR set-up, this study (P7) explored how specific VR features, such as environment, locomotion, auditory and visual presentation, could support a deeper understanding of pitch-colour cross-modal associations in a more immersive context. This study provided important insights that help answer the thesis question, highlighting design ideas that have the potential to support effective cross-modal associations between sound and colour. These insights include minimal environmental design to promote focus, adaptable locomotion options to reduce simulation sickness, and an expanded colour palette to offer users greater choice.

- **Chapter 6 (Study 3) - Exploring Cross-Modal Sound-Colour Associations for Music Learning in Immersive Virtual Reality**
  - RQ3: How do personalised & pre-defined pitch-colour cross-modal mappings compare in supporting learning outcomes in VR-based music exercises?

The final study (P8) provided an informed design to help facilitate pitch-colour cross-modal associations in an immersive VR environment. With the environment designed to help foster the cross-modal associations, a music test featuring eight lessons was created to assess how participants performed using pitch-colour cross-modal associations. In this study, the control group used pre-defined pitch-colour mappings (those where colours are already mapped to a specific note), and the experimental group formed their own personalised mappings (those where participants choose which colours are mapped to a specific note) in the VR application prior to the music test. The outcome of this study provides a basis for cross-modal design in VR environments and highlights a use case for the research in a music learning context.

1.4 Contribution to the Field

The research conducted as part of my PhD has made a design-oriented contribution by developing and evaluating approaches to pitch-colour cross-modal associations in immersive environments, while also exploring features that impact how the connections between the auditory and visual senses are understood. These contributions have been disseminated through peer-reviewed international conferences and journals, relevant to affective computing, human-computer interaction, virtual reality, and multi-modal interfaces. From the eight papers highlighted below (See the list of papers in Appendix E), P5, P7, and P8, contributed directly to Chapters 4, 5, and 6 respectively.

This research has advanced knowledge in the following key areas:

1. **Advancing the Study of Pitch-Colour Associations:** Before my PhD, an initial study investigated the strengthening of pitch-colour associations using a guitar for auditory stimuli and a standard desktop for visual stimuli (O'Toole et al., 2019). My first published work from the PhD, P1, expanded on O'Toole et al. (2019) by exploring sound-emotion associations in greater depth, particularly examining the roles of personality traits, learning styles, and musical experience. This work was presented at `International Conference on Multimodal Interaction (ICMI)` (Qualis A2), contributing to discussions on the psychological and cognitive mechanisms underlying pitch-colour and sound-emotion links. Additionally, P3 reviewed sound-colour cross-modal associations, investigating how age, gender, personality, mood, and emotion shape auditory-visual perception. The findings were presented at `ICMI` (Qualis A2), informing broader interdisciplinary discussions on multi-sensory cognition and individual differences in cross-modal processing.

2. **Investigating Virtual Reality as a Research Platform for Cross-Modal Studies:** A Doctoral Consortium paper (P4), presented at the `International Conference on Affective Computing and Intelligent Interaction (ACII)` (Qualis B1), was the first step in evaluating VR as a platform for conducting cross-modal pitch-colour association research. Later, P5 investigated how participants formed pitch-colour associations in a low-immersion VR setting versus a standard desktop setting, providing key insights into the viability of VR as a research tool for multisensory studies. Findings were presented at `The International Conference on Advanced Visual Interfaces (AVI)` (Qualis A2), contributing to ongoing discussions on the benefits and limitations of VR for perceptual and cognitive research.

3. **Demonstrating the Role of Immersion and Interaction in Strengthening Cross-Modal Links:** The study presented in P7 examined how different immersive features and environmental settings impact participants when making pitch-colour associations. This research identified key elements, such as locomotion methods, interaction techniques, and auditory-visual presentation styles, that contribute to more effective and intuitive cross-modal associations in VR. Findings were published in the `Virtual Reality Springer Journal` (Q1 Journal), where they contributed to discussions on user engagement and perceptual learning in immersive virtual environments.

4. **Exploring VR-Based Cross-Modal Mapping in a Music Learning Application:** Building on insights from P7, the `Chromaesthetica` VR application was developed to explore music learning through pre-defined and personalised pitch-colour mappings. The study presented in P8 has been submitted to the *Computers & Education: X Reality Journal*; The article introduced an application designed using cross-modal associations in immersive VR music education, offering new perspectives on adaptive learning strategies in virtual environments. A demonstration of `Chromaesthetica` was showcased at the `International Conference on Mobile and Ubiquitous Multimedia (MUM)` (P6), with a Qualis score of B2.

An additional paper (P2) investigated virtual agents and transactive memory systems (TMS), offering insights into collaborative cognitive models. While not directly linked to this thesis's primary focus, these findings provide a foundation for future research on virtual instructional systems and adaptive learning frameworks in VR. Findings were presented at the `International Conference on Affective Computing and Intelligent Interaction (ACII)` (Qualis B1), contributing to discussions on Artificial Intelligence (AI)-driven instructional design and cognitive collaboration in virtual environments.

These studies, particularly P5, P7, and P8, have contributed to broader academic discourse on multi-sensory learning, virtual reality applications, and human-computer interaction, shaping future research in immersive auditory-visual learning and VR-enhanced cognitive training.

1.5 Overview of the Thesis

This thesis explores how immersive virtual reality environments can enhance pitch-colour cross-modal associations, progressing through theoretical foundations, experimental studies, and practical applications. The research is presented across seven chapters, each contributing to the overall investigation.

Chapter Two - Literature Review: This chapter comprehensively reviews existing research on pitch-colour cross-modal associations, synaesthesia, virtual reality, multi-sensory learning, and cross-modality in VR. It explores theories, such as multi-sensory integration, embodiment, and cognitive load theory in virtual settings, establishing the theoretical background that informs the research studies.

Chapter Three - Methodology: This chapter details the methodological framework used to design and conduct the studies that address the overarching research question. It outlines the researcher's epistemological position, the research design and rationale, and the quantitative and qualitative methods employed in each study in this thesis.

Chapter Four - Sound and Colour: Evaluating Auditory-Visual Tests in Virtual Reality and Traditional Desktop Settings: This chapter (see P5) presents a study investigating pitch-colour consistency tests conducted in two different technological settings: a standard desktop environment and a VR HMD. The

study evaluates whether the testing medium influences participants' pitch-colour associations, providing insights into the viability of VR as a research tool for cross-modal studies.

Chapter Five - Investigating Musical Pitch-Colour Associations in Immersive Virtual Reality Environments: Building upon the previous study, this chapter (see P7) examines how immersive features in VR impact pitch-colour associations. Participants engage in two distinct VR environments, each incorporating different levels of auditory presentation, visual stimuli, locomotion methods, and environmental design. The findings provide insights into how the design of virtual environments can be optimised to support music learning through pitch-colour cross-modal associations.

Chapter Six - Exploring Cross-Modal Sound-Colour Associations for Music Learning in Immersive Virtual Reality: This chapter (see P8) focuses on the application of pre-defined versus personalised pitch-colour mappings in a VR-based music learning context. Using a mixed-methods approach, the study investigates how different mapping strategies impact engagement, learning effectiveness, and cognitive association when participants learn musical concepts using immersive VR.

Chapter Seven - Discussion and Conclusion: The final chapter synthesises the key findings from this research, examining the common themes that emerged across the three studies. It then highlights the practical and theoretical contributions of the thesis and connects these insights back to the central research question. The chapter also acknowledges the limitations of the studies and outlines potential directions for future research before concluding.

Chapter 2

Literature Review

This thesis seeks to understand the capabilities of immersive VR environments in facilitating cross-modal associations, specifically, the relationship between pitch in the auditory sense and colour in the visual sense. As outlined in the previous chapter, this research draws upon multiple fields of study, including HCI, psychology, and music learning, to investigate how Extended Reality (XR) technologies can support cross-modal associations. This chapter outlines the main research that informs the background and direction of the thesis. Chapters 4, 5, and 6 each contain additional background sections specific to the published studies they present.

The chapter begins by outlining how the literature was identified, selected, and synthesised, setting the foundation for the subsequent analysis. Next an examination of cross-modal associations and their close relationship to synaesthesia, focusing on pitch-colour associations and related auditory-visual mappings is investigated. It then explores VR as a medium for investigating cross-modal experiences, highlighting key immersive features and relevant design theories in multimodal learning. Finally, a review of prior research that intersects cross-modality and VR, and identifies key gaps in the existing literature that this thesis aims to address.

2.1 Literature Review Methodology

This literature review adopted a narrative interdisciplinary and iterative approach rather than a more formal systematic review. A more formal review was not adopted, as predefined keyword strategies and rigid inclusion criteria can limit the identification of relevant interdisciplinary work. An iterative strategy allowed for backward and forward citation tracking and the identification of relevant work across disciplinary boundaries that may not share consistent terminology. This approach helped to contextualise a broad question that spans multiple disciplines, which would be harder to achieve with a systematic review. Given the interdisciplinary nature of the research, which spans human-computer interaction (HCI), psychology, music, virtual reality, neuroscience, and education, a targeted, iterative search strategy was used. Searches were conducted primarily using `Google Scholar`, `ACM Digital Library`, `IEEE Xplore`, and `ResearchGate`. Studies were selected based on their relevance to the research questions, their contribution to understanding cross-modal associations, and their application in interactive and immersive environments. The selected literature was analysed collectively to identify recurring concepts, findings, and gaps across disciplines. These insights informed the development of the research

questions and shaped the design of the VR studies. This more flexible approach is in line with epistemological standpoint of the researcher and the mixed methods approaches used.

2.2 Cross-Modal Associations

When we talk about using our different senses or modalities together to form perceptions about what is happening in the world, the term multimodal is a more widely understood phrase. In HCI, a multimodal system processes two or more combined user input modes (such as speech, touch, gaze, head, and body movements) in a coordinated manner with a multimedia system output (Jacko and Sears, 2003). These different modes work together to construct a singular meaning in a given situation. Understanding multimodal design has become increasingly important in terms of multimedia, technology, and how we interact in digital environments. While the visual modality is the predominant source of information for humans, the other modes (such as the auditory, olfactory, and tactile) help create a full understanding of the world around us (Martin et al., 2022). VR is often described as an immersive, multimodal technology that delivers realistic virtual environments through a combination of visual, auditory, and haptic feedback. While the focus on multimodal design in HCI has provided improved ways of interacting with technology, another research field has been overlooked, which could provide a new level of interaction in virtual environments and a new level of perception for users. The research field is cross-modality. With cross-modal interactions, the sensory modalities relate to each other when interacted with in a multimodal environment, but these modalities also influence each other (Guo and Jiang, 2023). A deeper look at cross-modal associations is presented in the next section.

2.2.1 What are Cross-Modal Associations?

Cross-modal associations or correspondence refers to the systematic association or link between distinct sensory modalities, where seemingly unrelated stimuli, properties, or dimensions from different senses (such as sound, sight, or touch) connect in a predictable manner (Deroy and Spence, 2016; Sun et al., 2024). Examples of these sensory links can be found between loudness (in the auditory sense) and brightness (in the visual sense) (Marks, 1987).

Before continuing this section on cross-modal associations, it is useful to clarify the terminology used throughout the thesis when referring to connections between sensory modalities. Terms such as cross-modal associations, cross-modal correspondences, cross-modal design, cross-modal interactions, cross-modal perception, cross-modal sensitivity, and cross-modality are used at various points in this work. Some of these terms are used interchangeably; for instance, cross-modal associations and cross-modal correspondences both describe the perceived compatibility between stimulus features across different sensory systems (Spence, 2011). Others are more context-specific: cross-modal design refers to the intentional application of cross-modal principles in system or environment design. These distinctions help clarify how cross-modality is conceptualised within different parts of the thesis.

Spence (2011) presents three principal types of cross-modal correspondence: statistical, structural, and semantic. These different types of correspondence have de-

veloped differently, where one is a natural correlation (statistical), another can occur due to neural connections formed since birth (structural), and the last type occurs when stimuli in two dimensions overlap (semantic) (Spence, 2011).

Statistical

Statistical cross-modal correspondences occur when our brain adapts to the regular patterns or correlations found in the natural world. Correspondences of this nature would be between pitch-elevation, pitch-size or loudness-size, where, for example, certain physical properties, like the sound an object makes when struck, are consistently linked to the size or shape of that object. Our brain internalises these correlations through experience, forming associations between sensory stimuli across different modalities (i.e. sound and sight). This is explained using Bayesian integration theory, which suggests that our brain uses prior knowledge (or *"priors"*) to predict and interpret sensory information based on learned regularities (Spence, 2011, (p. 984)).

Structural

Structural cross-modal correspondences arise from the way our brain is wired to process sensory information. They occur because different senses share similar neural mechanisms. For example, our brain responds with increased neural activity as a stimulus becomes more intense, whether light or sound. This may explain why we often link loud sounds with bright lights. Additionally, certain areas of the brain, like those handling magnitude, might process similar features across senses, leading to natural associations between them. Essentially, these correspondences are a by-product of the brain's structure and its efficient way of processing different sensory inputs (Spence, 2011)

Semantic

Semantically mediated cross-modal correspondences occur when we use the same linguistic terms to describe stimuli across different senses. For example, the terms *"high"* and *"low"* are commonly used to describe both pitch in sound and elevation in space. This type of correspondence is shaped by language and only develops after we acquire it. Unlike other types of cross-modal correspondences, semantic associations are context-dependent and do not rely on inherent sensory similarities; rather, they are based on the shared descriptive words used across different modalities. These associations typically influence decision-making rather than perception itself (Spence, 2011).

Examining cross-modal associations through the previously discussed semantic, structural, and statistical types helps us understand how different sensory features are connected. However, not all associations fit neatly into one category. In the case of pitch-colour associations, for example, it is not always clear whether they are semantic, structural, or statistical. As Spence (2011) points out, even widely reported links, like pitch and brightness, are difficult to pin down to a single type. This suggests that a mix of different factors may shape pitch-colour associations. In the context of this thesis, these associations are not treated as strictly semantic,

structural, or statistical. Instead, they are influenced by elements of all three—especially when explored in immersive VR environments where multiple sensory cues and immersive features are presented together.

2.2.2 Auditory-Visual Cross-Modal Associations

Cross-modal associations between the auditory and visual senses have been extensively researched. Previous research has demonstrated various cross-modal correspondences between auditory and visual features. Table 2.1 originally presented in Spence (2011) shows the different visual features and their congruent auditory feature from different studies in the area. These include congruent associations between pitch-size, pitch-elevation, pitch-lightness, pitch-brightness and pitch-direction of movement. While Spence (2011) has shown how some of these different cross-modal correspondences are understood by the inner-workings of the brain, such as pitch-size using statistical mediated associations and pitch-elevation using semantically mediated cross-modal correspondence, some connections, like pitch-brightness, have been hard to pinpoint how and where it is processed in the brain.

Focusing on sound-colour associations and cross-modal correspondences, a few studies have presented congruent associations between pitch and lightness, brightness, and hue. Ward et al. (2006), found that both subject types—synaesthetes and controls (non-synaesthetes)—used similar cross-modal mechanisms when associating pitch with lightness in the visual sense, but at different consistency levels. In terms of brightness, it has been shown that higher pitch sounds find a congruent association with higher brightness and lower pitch with lower brightness (Calvert et al., 2004).

The colour feature of hue has also been studied in relation to its interaction with cross-modal correspondence, yielding varying results. One study measured Reaction Times (RTS) of participants, identifying red versus blue targets under different auditory conditions. It found that the presence of a non-informative tone reduced overall RT. However, there was no selective effect of tone frequency on colour detection, suggesting that the stage of information processing for colour selection differs from that for location selection (Bernstein et al., 1971).

The emotion mediation hypothesis proposed by Palmer et al. (2013) suggests that emotional responses to music influence how people form associations between sound and colour. Their findings showed that participants tended to choose colours that reflected the emotional character of the music they were hearing. Similarly, in the case of single musical notes, O'Toole et al. (2019) found that participants matched emotions and colours to the pitch being played. For example, the lowest pitch in the scale was associated with dark blue and sadness, while higher pitches were linked to brighter colours like green or cyan and emotions such as happiness.

In summary, while there is a growing body of research exploring auditory-visual cross-modal associations, especially between pitch and colour, it is clear that these links are not always straightforward. Different auditory features, such as pitch, timbre, or rhythm, tend to associate more naturally with different types of visual features, such as pitch with colour or brightness, and timbre with shape or texture. Additionally, emotional content can significantly influence these associations, with users often selecting colours or visuals that match how the sound makes them feel. This highlights that no universal pathway exists for how sound maps onto vision. In-

Table 2.1: Summary of cross-modal correspondences that influence performance on the speeded classification task with null results (adapted from Spence (2011)).

Auditory	**Visual**	**Cross-modal**	**High-Pitch / Loud**	**Studies**
Pitch	Elevation	Yes	High elevation	Ben-Artzi and Marks (1995); Bernstein and Edelstein (1971); Evans and Treisman (2010); Melara and O'Brien (1987); Patching and Quinlan (2002)
	Brightness	Yes	Brighter stimulus	Marks (1987)
	Lightness	Yes	Lighter stimulus	Martino and Marks (1999); Melara (1989); Marks (1987)
	Shape/angularity	Yes	More angular shape	Marks (1987)
	Size	Yes	Smaller object	Evans and Treisman (2010); Gallace and Spence (2006)
	Spatial frequency	Yes	High spatial frequency	Evans and Treisman (2010)
	Direction movement	Yes	Upward movement	Clark and Brownell (1976)
	Contrast	No	N/A	Evans and Treisman (2010)
	Hue	No	N/A	Bernstein et al. (1971)
Loudness	Brightness	Yes	Brighter stimulus	Marks (1987)
	Lightness	No	N/A	Marks (1987)

stead, multiple factors, from individual perception to emotional context and sensory properties, can influence how these associations are formed and experienced.

2.2.3 Relationship with Synaesthesia

As highlighted in Section 1.1, there is a close relationship between the study of cross-modal associations and synaesthesia. Martino and Marks (2001) introduced the distinction between *"strong"* and *"weak"* synaesthesia. In this work, they characterised strong synaesthesia as the *"vivid imagery in one modality in response to stimulation in another one"*. In contrast, weak synaesthesia is described as *"cross-sensory correspondences expressed through language, perceptual similarity, and perceptual interactions during information processing"*. From this perspective, strong synaesthesia aligns with traditional definitions of the condition, such as that proposed by Cytowic (1989), while weak synaesthesia corresponds more closely to cross-modal correspondences or associations, described in Section 2.2.

Other researchers have explored similar questions regarding the overlap between synaesthesia and cross-modal associations. For example, Campen (2010) conducted auditory-visual consistency tests with art students, finding a wide distribution of scores. Some participants exhibited consistency levels closer to those of individuals with synaesthesia, while others displayed a broader range. This spectrum of responses supports the strong-weak synaesthesia continuum proposed by Martino and Marks (2001).

Figure 2.1 illustrates the proposal of synaesthesia as a continuum, presenting previous research on the topic and indicating how it informs the positioning of this study along that spectrum.

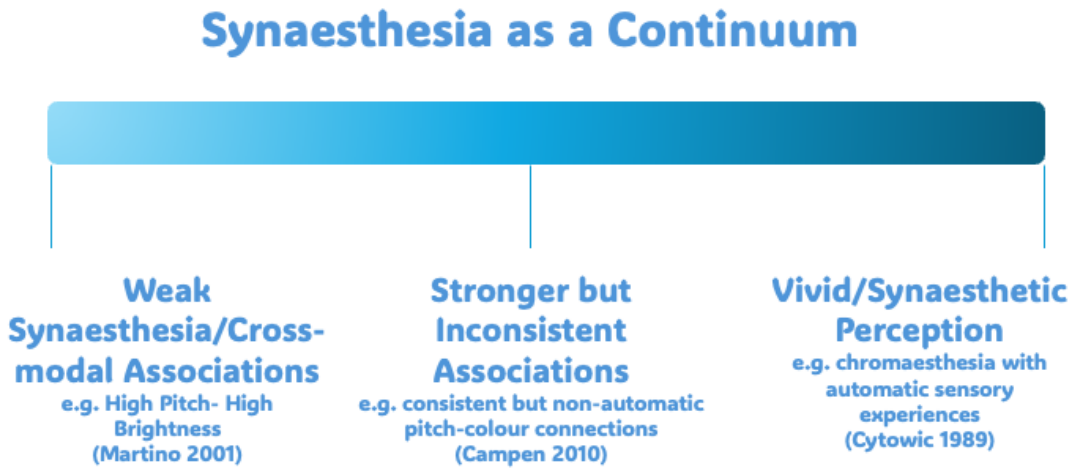


Figure 2.1: Synaesthesia conceptualised as a continuum, ranging from cross-modal associations (*"weak"* synaesthesia) to full synaesthesia with highly consistent and automatic experiences (Martino and Marks, 2001; Campen, 2010; Cytowic, 1989).

While many researchers acknowledge a conceptual link between these two perceptual phenomena, others argue that they are empirically distinct, sharing only

superficial similarities (Parise and Spence, 2012). However, from a neuroscientific perspective, Ramachandran and Hubbard (2003) suggests that both synaesthesia and cross-modal associations may involve similar neural mechanisms. They proposed that the Temporo-Parietal-Occipital (TPO) junction, a brain region where sensory modalities converge, is involved in both types of processing. Supporting this, Ward et al. (2006) demonstrated that sound-colour synaesthesia relies on the same perceptual mechanisms as those found in typical cross-modal processing.

Together, these findings suggest that even if synaesthesia and cross-modal associations are not identical phenomena, they are intimately related. Whether synaesthesia is viewed as a more vivid or automatic manifestation of cross-modal processing, the two fields have informed each other substantially. Research into synaesthesia has provided valuable insights into the mechanisms of cross-modal perception, and vice versa, contributing to a more comprehensive understanding of sensory integration.

2.3 Synaesthesia

Synaesthesia can be defined as a union of the senses, with the term derived from the Greek words *"syn"* (union) and *"aisthesis"* (sensation) (Cytowic, 2012). It arises when stimulation of one sense (the inducer) triggers a reaction in an unstimulated secondary sense (a concurrent) (Cytowic, 2003; Curwen, 2018). It is a rare condition that affects about 2-4% of the general population. Synaesthesia occurs automatically and is generally consistent over time (Curwen, 2018; Farina et al., 2017). So far, more than 75 different types of synaesthesia have been found, the most prevalent being *"graphemes to vision"* synaesthesia (1 in 44 people) (Day, 2022). Other prevalent types include *"general sounds to vision"* and *"musical sounds to vision"*, with 16.21% and 18.05% of synaesthetes (from a database of 1297 individual synaesthetes) respectively identifying as having these types (Day, 2022).

Synaesthesia has long been a challenging phenomenon to understand. Much of the early research focused on cognitive neuroscience, aiming to identify which areas of the brain are responsible for triggering synaesthetic experiences (Ramachandran and Hubbard, 2001, 2003; Brang and Ramachandran, 2011). Additionally, it has been investigated through other research domains, such as psychology (Curwen et al., 2023; Asano and Yokosawa, 2011; Janik McErlean and Banissy, 2017; Spector and Maurer, 2009), art, music (Rudenko and Cabral, 2019; Shirokova, 2021; Chen and Tsao, 2023; Haverkamp, 2022), philosophy (Macpherson, 2007; Brogaard, 2020), and design (Merter, 2017; Wright and Kefalidou, 2021; Lee et al., 2019).

The following sections explore the different ways that synaesthesia can be classified—not just by type, such as sound-colour, but also by how it is experienced and develops. Individuals may differ in whether their synaesthesia is developmental or acquired, and whether they perceive it internally or externally. These overlapping dimensions highlight the complexity of synaesthetic experiences beyond simple sensory pairings.

2.3.1 Differences in Synaesthesia

Synaesthesia can be classified not only by the type of sensory pairing (i.e. pitch and colour), but also by how it develops and how individuals experience it. The most widely recognised form is developmental synaesthesia. Hill (2006) explains that it

typically emerges in newborns and is marked by vivid, automatic, involuntary, and unlearned sensory experiences. At four to six months old, the neural pathways that support synaesthesia usually disconnect, with around one in two thousand people retaining the connection into adulthood (Ward et al., 2006; Goller et al., 2009).

Other synaesthetic experiences are acquired and are not present at birth but emerge after life-changing events such as head injuries, tumours, or physiological conditions (Hill, 2006; Goller et al., 2009). There have also been cases of drug-induced synaesthesia which produce short-lived, often inconsistent experiences during altered states of consciousness (Hill, 2006). This supports the theory that synaesthesia is an exaggeration of existing neural pathways rather than the creation of entirely new ones (Ward et al., 2006).

A further distinction is metaphorical synaesthesia, where language is used to describe one sense using another, such as *"the blues"* or *"sticky purple-grey smell"* (Hill, 2006; Tsang and Schloss, 2010). While common in everyday speech, individuals with synaesthesia tend to have a more refined and consistent vocabulary for describing such experiences (Curwen, 2018).

Going beyond how synaesthesia develops in an individual, there is a phenomenological difference in how it can be experienced. Associators synaesthetes report colours or shapes *"in the mind's eye"*, while projectors experience them as externally located in space (Curwen, 2018). These are not just subjective differences, as studies suggest they may involve distinct cognitive mechanisms. Another proposed distinction is between *"higher"* and *"lower"* synaesthetes: the former are thought to process at a conceptual level, while the latter operate at an earlier perceptual stage (Curwen, 2018).

Taken together, these differences show that synaesthesia is not one fixed condition but is experienced by its type (e.g. chromaesthesia), its beginnings (e.g. developmental), how it is perceived (e.g. projectors), and by its cognitive mechanisms (e.g. higher or conceptual level). These differences are important when thinking about cross-modal associations and having synaesthesia research inform the design of VR environments that aim to support or strengthen pitch and colour cross-modality. Whether trying to replicate synaesthetic experiences or make these connections feel more intuitive and engaging, it is important to understand that people respond to these pairings in different ways. This understanding can help shape better design choices for creating meaningful multisensory experiences in VR.

2.3.2 Chromaesthesia or Sound-Colour Synaesthesia

Synaesthesia experienced between sounds and colours is called chromaesthesia or sound-colour synaesthesia, and is found in an estimated 0.2% of the general population (Farina et al., 2017). A synaesthete will associate different types of sounds with colour in the visual sense, such as musical notes, musical sounds and general sounds (Day, 2022). The reverse of sound-colour synaesthesia, where the colour is the inducer and the sound is the concurrent, has also been observed; however, this is much rarer (Goller et al., 2009). Within the context of musical sounds, pitch can be the most prominent feature that informs associations with the visual sense; however, other sound features need to be considered, such as loudness, tempo, timbre, tonality, consonance/dissonance, pitch class and emotional response (Ward et al., 2006). These other auditory features also contribute to the visual experience that

someone with chromaesthesia might perceive. However, some features may rely on normal cross-modal associations alongside chromaesthesia to create the full visual picture.

A book by Campen (2010) that explores synaesthesia in art and science reveals interesting insights from a synaesthetes' experience listening to music. An excerpt from the person with chromaesthesia describing their experience listening to a Prince song presents how the different musical features are perceived visually, *"The lines produced by guitar sounds are jagged and blue in colour. The higher the pitch, the lighter the colour, tending towards white. Also, these lines disappear left to right from the image. The more strings sound together, the more jagged the lines that I see becomes"* (Campen, 2010, (p. 12)). This description shows both synaesthetic elements and normal cross-modal associations at play in the person's musical experience. As highlighted in Section 2.2.2, research has found that people generally perceive higher-pitched sounds as lighter in colour (Curwen, 2018; Ward et al., 2006; Goller et al., 2009). In addition, studies on timbre have shown that harsh instrument timbres are often associated with angular shapes (Adeli et al., 2014). Past research into synaesthesia has shown some similarities with cross-modal associations found in the general population (Ward et al., 2006; Williams et al., 2015; Lacey et al., 2016). Given these overlaps—along with the complexities of synaesthesia in terms of type, category, and classification—it is possible that the two areas of study share a closer relationship than previously assumed.

With the main focus of this thesis being pitch-colour cross-modal associations, it is important to highlight the discussion of chromaesthesia presented above. As shown, there are similarities between sound-colour cross-modal associations and chromaesthesia. If these are considered as two ends of the same spectrum (as illustrated in Figure 2.1), then research into chromaesthesia can help inform the central research question.

Exploring the relationship between pitch and colour using VR technologies also provides an opportunity to design synaesthesia-like experiences and investigate how they might impact non-synaesthetes when forming pitch-colour associations. One area that could benefit from further development is synaesthesia testing, particularly by moving beyond traditional pen-and-paper or desktop-based methods toward more immersive formats. The next section will look at the different tests available, how they differ, and how they have progressed over time.

2.3.3 Testing for Synaesthesia

As this thesis centres on pitch-colour cross-modal associations, this section focuses on synaesthesia tests that specifically investigate sound-colour links. Previous research on synaesthesia has helped develop tests for determining whether an individual is a synaesthete or a non-synaesthete. While the discussion in Section 2.3.1 on external (Projectors) versus internal (Associators) perception highlights the complexity of synaesthetic experience, it also underscores the difficulty of designing effective testing methods. Due to its complex and varied ways synaesthesia is experienced, several tests have been developed to measure the consistency and nature of these associations (Curwen, 2018). These tools are useful not only for identifying synaesthetes but also for revealing a broader spectrum of cross-modal sensitivity. For example, as previously mentioned in Section 2.2.3, Campen (2010) used such tests with art

students and found a wide range of scores—from levels consistent with synaesthesia to much weaker sensory associations—highlighting how these assessments can help contextualise where an individual's experience sits along a continuum. These tests usually involve a type of consistency or speed classification test to assess the person's certainty in their associations between two stimuli. These tests do not accommodate all types of synaesthesia and are usually limited to the more common forms observed in the general population.

The next section presents an overview of the most prevalent synaesthesia tests, focusing on the `The Synaesthesia Battery` (Eagleman et al., 2007), `Test of Genuineness (TOG)` (Asher et al., 2006; Johnson et al., 2013), `Netherlands Colour Synaesthesia (NeCoSyn)` method (Campen and Froger, 2003), and the `Implicit Association Test (IAT)` (Parise and Spence, 2012; Lacey et al., 2016).

Synaesthesia Battery:

`The Synesthesia Battery`[1] is an online platform that can test multiple types of synaesthesia. It has been independently tested and validated, and is one of the most used tests in synaesthesia research to date (Carmichael et al., 2015). This browser-based testing platform offers separate portals for participants and researchers, making it easy to track each participant's data through the research portal. It features a straightforward interface for conducting synaesthesia tests and provides multiple options for exporting the collected data. In the battery tests, over twenty synaesthesia tests are available, ranging from musical pitch-colour tests to vision-smell synaesthesia test (Eagleman et al., 2007).

This section focuses on the musical pitch to colour test available in the `The Synesthesia Battery`. The test involved hearing thirteen different musical pitches, ranging from middle C (C4 - 261.63 Hz) to one octave above (C5 - 523.25 Hz). These thirteen pitches are played three times each at random to make up thirty-nine auditory stimuli that need to be associated with a colour from a colour picker chart. This test is scored on how close the three colour choices for each pitch are, with a score of under 1.0 deemed to be the level of a synaesthete and above 1.0 the level of a non-synaesthete (Eagleman et al., 2007). This score classification has been challenged, with research suggesting that the threshold score of 1.0 for identifying synaesthetes may be overly conservative, potentially excluding individuals who still show meaningful consistency in their associations. It has been proposed that a score can be set marginally higher to be classed as a synaesthete (Carmichael et al., 2015).

While the `The Synesthesia Battery` tests main research purpose is to help identify certain synaesthetic abilities in participants, it has also helped understand how non-synaesthetes associate between auditory-visual senses, specifically in terms of musical pitch and colour. Campen (2010), in their study involving 223 fine art students, comprising both synaesthetes and non-synaesthetes, anticipated observing two distinct peaks in their data. They expected the mean synaesthesia scores of synaesthetes to cluster at a consistently high level (approximately 90%) and those of non-synaesthetes to fall at a significantly lower level (around 40%). However, their findings revealed a more continuous distribution of synaesthetic experiences across the entire range of 0-100%, with no clear distinction between synaesthetic and non-synaesthetic participants. This unexpected result prompted them to re-

[1] `https://synesthete.ircn.jp/home`

vise their original research question to a more nuanced one: *"How strong is your synaesthesia?"* (Campen, 2010).

Such findings align with other research that challenges traditional binary classifications of synaesthesia, suggesting instead that synaesthetic experiences may exist on a spectrum, ranging from strong synaesthetic perception to weak synaesthesia.

Test of Genuineness (TOG-R):

The `Test of Genuineness (TOG)` was created in 1987 by Baron-Cohen et al. (1987) and has been improved over the years, becoming the *"gold standard"* for assessing the consistency of synaesthetes' reports (Asher et al., 2006; Johnson et al., 2013). Baron-Cohen et al. made improvements to the TOG throughout the years, as documented in their papers from 1987 (Baron-Cohen et al., 1987), 1993 (Baron-Cohen et al., 1993), and 1996 (Baron-Cohen et al., 1996); however, these tests did not include assessments involving the auditory sense.

A revised version of the `Test of Genuineness (TOG-R)` was created, featuring a sound-colour test alongside a more refined scoring system (Johnson et al., 2013). The sound stimuli for the TOG-R included forty-eight sounds, such as instrumental, naturally occurring, artificial environmental sounds, and vocal exclamations. The `Cambridge Synaesthesia` charts were used to represent colours, consisting of two hundred and thirty-eight numbered colours. A consistency of 71.3% for synaesthetes and 33% for non-synaesthetes was found, suggesting that this test can accurately distinguish between the two groups (Johnson et al., 2013). The TOG-R appears to have a similar testing method to `The Synaesthesia Battery`. However, Asher et al. (2006) added a new dimension to the colour synaesthesia test that would categorise the synaesthetes into *"broad-band"* (scoring 80%-100%), and *"narrow-band"* (scoring 37%-71%) synaesthetes (Johnson et al., 2013). This categorisation seems to consider the non-binary view of synaesthesia, at least in terms of sound-colour associations, that was presented by Campen in Section 2.3.3.

NeCoSyn:

The `NeCoSyn` test, developed by Campen and Froger (2003), aims to explore individual profiles of colour synaesthesia. Drawing from the `Swedish Natural Colour System` and the `TOG` for colour-word synaesthetes, this test offers personalised synaesthetic profiles in dimensions of hue, chroma, and blackness (Campen and Froger, 2003). The `NeCoSyn` test is capable of assessing associations between colour-word, colour-taste, colour-odour, and colour-music.

In this study, Campen and Froger (2003) refers to the general population or *"normal"* participants as having general synaesthesia, while referring to synaesthetes as having *"particular synaesthesia"*. In a book by Campen (2010), studies show that the `NeCoSyn` test identified more participants as synaesthetes than the TOG test. The `NeCoSyn` test offers interesting insights into specific dimensions of synaesthesia and how they relate to both general and particular types of synaesthetes. The hue dimension was found to be related to *"particular synaesthesia"*, while the dimension of blackness was related to general synaesthesia. Also, for the colour-music synaesthesia test, chroma displayed higher consistency than blackness and hue (Campen and Froger, 2003).

A study by Sun et al. (2018) on cross-modal correspondence found that the reaction times for sound-lightness were quicker than those for sound-hue when testing non-synaesthetes on different auditory properties with colour. It might be that non-synaesthetes or general synaesthetes are more in-tune to lightness or blackness, while synaesthetes are better at perceiving hues regarding associations with sound stimuli.

Implicit Association Test (IAT):

The `Implicit Association Test (IAT)` measures the strength of associations between a pair of concepts (Parise and Spence, 2012). Previous research has used it to understand synaesthesia and normal cross-modal associations (Lacey et al., 2016). Parise and Spence (2012) used a modified `IAT` test to replicate classic cross-modal association tests, also highlighting two new correspondences. Their study included the takete-maluma and mil-mal tests—nonsense-word tasks that measure shape-sound associations, where participants tend to match sharp-sounding words (like *"takete"* or *"mil"*) with angular shapes, and softer-sounding words (like *"maluma"* or *"mal"*) with rounded shapes. In addition, the study examined audio pitch-size pairings and revealed new correspondences between pitch and visual features such as angularity and object size (Parise and Spence, 2012).

A study by Lacey et al. (2016) used the `IAT` to explore auditory pitch-visual elevation, auditory pitch-visual size, and auditory pseudo words-visual shapes associations to compare synaesthetes' and non-synaesthetes' sensitivity to cross-modal associations. Results showed that synaesthesia heightens universally experienced cross-modal correspondences, but only when these involve sound symbolism. This result only partly confirms the idea of a continuum between synaesthesia and cross-modal correspondences (Lacey et al., 2016).

To summarise this section on synaesthesia, it is clear that the condition is both complex and varied, with different types, such as sound-colour synaesthesia, being experienced and acquired in different ways. These differences make synaesthesia a difficult phenomenon to define, particularly when it comes to how it is perceived and processed in the brain. While not all forms of synaesthesia are thought to exist on a continuum, this idea has been proposed for sound-colour synaesthesia, which informs the focus of this thesis: pitch-colour cross-modal associations.

By viewing pitch-colour associations and chromaesthesia as part of a shared spectrum, we can begin to consider how these sensory connections might be supported or strengthened—even in non-synaesthetes. Virtual reality offers a promising way to explore this. As a multisensory and interactive platform, VR can go beyond traditional testing approaches (as discussed in Section 2.3.3) by offering more engaging and immersive environments that may help users form stronger cross-modal links.

The next section will explore the key features of VR, including immersion, presence, interaction, and spatial audio, and consider how these features might support the formation and reinforcement of pitch-colour associations, particularly in individuals without natural synaesthetic tendencies. This will help inform the central research question and identify potential design strategies for enhancing cross-modal perception through VR.

2.4 VR Technologies and Capabilities

Virtual Reality (VR), as a subset of Extended Reality (XR), refers to computer-generated simulations to immerse users in interactive, three-dimensional environments. XR is an umbrella term encompassing Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), each offering varying degrees of digital interaction with the physical world (Vasarainen et al., 2021). In the context of this research, VR is the primary focus due to its unique ability to fully immerse users in controlled, multisensory environments.

Historically, VR has evolved from early conceptual models in the 1960s, such as Morton Heilig's `Sensorama` and Ivan Sutherland's `Sword of Damocles` (Tutescu, 2022), to modern, widely accessible consumer headsets like the `Meta Quest` or `HTC Vive`. What once was a niche technological experiment has now become a mainstream platform for gaming, training, therapy, and education.

VR can be presented in various forms, commonly categorised as non-immersive, semi-immersive, or fully immersive systems. Non-immersive VR typically involves desktop-based or screen-based applications where users engage with virtual content via traditional displays without sensory encapsulation. Semi-immersive systems may use larger projection setups or multi-screen environments to provide partial immersion, often incorporating some degree of depth perception and spatial interaction. Fully immersive VR includes technologies such as head-mounted displays (HMDs) and Cave Automatic Virtual Environment (CAVE) systems, which surround the user with projected or rendered imagery on multiple surfaces to create a compelling sense of presence (Salatino et al., 2023). These systems aim to provide a first-person, embodied experience within a digitally rendered environment, often enhanced through high-resolution visuals and responsive user input, without necessarily relying on motion-tracked controllers or spatial tracking in all cases. The virtual environments produced can range from abstract representations to highly detailed, photorealistic simulations, offering various levels of interactivity and sensory realism depending on the system's configuration (Brooks, 1999).

With this thesis focusing on VR—predominantly using HMDs—to explore new understandings of cross-modality, it is necessary to examine the unique features that make VR a viable platform for investigating cross-modal associations of pitch-colour. Key capabilities that have positioned VR as a viable medium for training and learning include immersion, presence, interactivity, and spatial audio. The following sections will explore these features and examine how they contribute to the design and experience of cross-modal systems, particularly in the context of pitch-colour associations.

2.4.1 Immersion

Immersion is frequently described as the extent to which a VE delivers a comprehensive, realistic, or vivid sensory experience to a user. In a broad sense, it pertains to the objective and measurable characteristics of a system's capacity to simulate real-world cues (Slater and Wilbur, 1997; Jerald, 2015). Common factors that contribute to an immersive system include stereoscopic visuals, wider Fields of View (FOV) and Fields of Regard (FOR), and accurate head-based rendering, among others (Bowman and McMahan, 2007; Jerald, 2015). However, the term *"immersion"* is

used variably in the literature, with *"technical immersion"* typically describing system features such as those listed in the previous sentence. Immersion or immersive experiences have also been categorised in a more broad sense, with the term often being used interchangeably with the term *"presence"* (Bowman and McMahan, 2007).

A deeper explanation of immersive experiences that encompasses more than technological fidelity alone has been proposed. A framework by Balcerak Jackson and Balcerak Jackson (2024) defines immersion through four interrelated dimensions: *"representational immersion"*, concerning the realism and coherence of the environment; *"participatory immersion"*, referring to the user's ability to interact with and affect the virtual world; *"affective immersion"*, involving emotional engagement; and *"narrative immersion"*, describing absorption in unfolding events or storylines. Together, these dimensions frame immersion not as a purely technical measure but as a multi-faceted experience shaped by emotional, cognitive, and interactive factors that incorporate elements of presence and other features of VR.

While there is confusion within the research area about immersion and how it relates to other features that help create *"immersive experiences"*, many adopt the conceptual distinction by Slater and Wilbur (1997):

- Immersion: An objective measure of the fidelity and inclusivity of the sensory stimuli provided by the technology.
- Presence: A user's subjective psychological response or state of *"being there"* inside the VE.

From this perspective, immersion is determined by the technical capacity of a system (for example, how wide the field of view is or how accurate the motion-tracking system is), while presence is influenced by the user's mindset, prior experiences, and context of use (Jerald, 2015). Indeed, individuals may exhibit different levels of presence even when using the exact same immersive system (Slater and Wilbur, 1997). While a certain level of immersion is important for creating a sense of presence in virtual reality, it is not the only factor; how present someone feels also depends on their emotions, attention, and the overall experience (Slater and Wilbur, 1997; Jerald, 2015).

While studies have shown a correlation between higher levels of immersion and increased presence (Slater and Wilbur, 1997), other studies have shown that *"full"* immersion is not always necessary (Bowman and McMahan, 2007; Grimshaw-Aagaard and Walther-Hansen, 2024; Cummings and Bailenson, 2016). Research by Cummings and Bailenson (2016) synthesised decades of empirical research (115 effect sizes from 83 studies) to understand whether enhancements in immersive system technology indeed produce corresponding increases in users' sense of presence. Although the aim of the review was to measure presence, it underscores which features of *"technological immersion"* most strongly contribute to a heightened feeling of *"being there"*. The results indicated an overall moderate effect; however, certain technical enhancements, such as user tracking, stereoscopic visuals, and wide fields of view, had a more substantial impact compared to other improvements like increased resolution or higher audio fidelity. While the paper focuses on presence as the outcome variable, it reinforces the notion that immersion (as an objective, quantifiable system feature) can meaningfully affect user experience.

In contrast, another study separated the effects of the level of immersion and 3D interaction technique on task performance within a VE. Two components of immersion—stereoscopy and field of regard—were manipulated, as were three 3D interaction techniques (`HOMER`, `Go-Go`, and the mouse-keyboard-based `DO-IT`) (McMahan et al., 2006). Interestingly, while the interaction technique significantly affected object manipulation time, the two immersion components did not show a statistically significant effect on performance for that particular task. These findings illustrate that *"full"* or *"high-end"* immersion is not always a prerequisite for effective user performance or experience. In specific contexts, other factors, such as usability, task design, or interaction style, may be more critical than strictly technical enhancements of immersion, when considering user performance and experience (Bowman et al., 2002).

From an educational perspective, research has found that immersion can enhance learning outcomes by facilitating multiple perspectives, enabling situated learning, and improving knowledge transfer (Dede, 2009). Immersive experiences can foster stronger engagement, potentially leading to deeper comprehension and more lasting impacts. Nonetheless, some scholars caution that not all learning scenarios demand the highest levels of immersive fidelity (Young et al., 2025; Makransky et al., 2019). A study by Young et al. (2025) found that the XR learning scenario was significantly higher for Stimulation and Novelty (items on the User Experience Questionnaire). However, participants did not significantly enhance their perceived learning. Similarly, Makransky et al. (2019) found that while learners reported higher presence and motivation in high-immersion VR simulations, they demonstrated lower learning outcomes and exhibited greater cognitive load. Practical considerations, such as cost, technical complexity, and the instructional method, must be weighed against the potential incremental benefits of ever-greater immersion (Dede, 2009; Young et al., 2025).

In summary, immersion is a crucial concept in virtual reality research, often intertwined with the concept of presence. On the one hand, immersion can be viewed objectively, focusing on measurable technological characteristics such as display fidelity, field of view, or interaction tracking. On the other hand, some scholars argue for more nuanced perspectives, suggesting that immersion spans not only technical realism but also user interactivity, emotional investment, and narrative engagement. Previous research demonstrates that greater technical immersion does not universally guarantee enhanced user experience or performance, highlighting that other design factors and user-centred considerations are equally important. This nuance sets up the next discussion on presence, highlighting that while immersion can help create presence, it is not the only factor—presence also depends on the user's subjective response and the context in which the VR is experienced.

2.4.2 Presence

As discussed in Section 2.4.1 on Immersion, VR experiences depend not only on technological fidelity (e.g., high-resolution displays, wide fields of view) but also on a user's subjective sense of *"being there"* in a VE. This user-centred phenomenon is typically referred to as presence (Slater and Wilbur, 1997). Although the literature uses terms such as telepresence, mediated presence, co-presence, and virtual presence (Lee, 2004), presence remains the most commonly accepted and easily understood

label (Lee, 2004). However, the concept is far from straightforward, as scholars often define and categorise it through different types, scales, and dimensions.

Work by Lombard and Jones (2015) identified seven forms of presence: spatial, social, self-engagement, realism, cultural, and parapresence. In the context of VR, spatial presence, the sensation of physically occupying virtual space, is especially relevant (Felton and Jackson, 2022). Schubert et al. (2001) presented a three-dimensional model that similarly described presence in VEs as comprising spatial presence (feeling located within the environment), involvement (psychological engagement with tasks or narratives), and realism (perceptions of authenticity or credibility). These categories underscore that presence is multidimensional; it arises from how users cognitively, emotionally, and behaviourally interact with the immersive system.

Because presence is subjective, researchers have developed various scales and questionnaires to assess it. Key instruments include:

- Witmer and Singer Questionnaire (WS): A relatively lengthy (32-item) measure examining multiple presence dimensions, such as involvement and sensory fidelity (Witmer and Singer, 1998).
- Slater-Usoh-Steed Questionnaire (SUS): A shorter, 6-item measure that primarily focuses on spatial presence (Usoh et al., 2000).
- iGroup Presence Questionnaire (IPQ): A 14-item instrument covering spatial presence, involvement, and realism (Schubert et al., 2001).

Each scale offers benefits and drawbacks, for instance, the SUS is concise but narrower in scope, while the WS and IPQ examine broader facets of presence. Ultimately, choosing which to administer often depends on the study's time constraints and the specific dimensions of presence one aims to measure.

Studies repeatedly show that interaction and user agency strongly impact presence. In one investigation, self-reported *"interaction possibilities"* significantly correlated with spatial presence but less so with involvement or realism (Regenbrecht and Schubert, 2002). Notably, even an illusion of interaction—where users believed they could affect the environment, but no actual interaction occurred—boosted spatial presence (Regenbrecht and Schubert, 2002). These findings suggest that interactive design choices in VR (e.g., how users manipulate objects or navigate environments) can bolster the user's sense of *"being there"*, even if the tasks themselves are relatively simple.

Beyond spatial presence, VR can evoke heightened affective responses. Riva et al. (2007) found a circular relationship where emotionally charged environments increased presence and higher presence amplified emotional states. Similarly, Pallavicini and Pepe (2019) observed that a VR racing game elicited more positive emotions than a desktop version, likely due to higher presence. This connection between presence and emotional engagement has important implications for educational or therapeutic VR applications, where emotional resonance can facilitate deeper learning or psychological impact (Riva et al., 2007; Pallavicini and Pepe, 2019; Dede, 2009).

Multimodal integration in VEs has also been shown to enhance presence. Marucci et al. (2021) found that trimodal (visual-auditory-tactile) stimulation enhanced users' subjective sense of presence and improved cognitive performance under

high perceptual load. Glenn and Coxon (2024) demonstrated that individual differences in multisensory processing help predict variations in presence, suggesting that people's innate or learned skills for combining sensory information influence how deeply they feel *"inside"* a virtual environment. These findings suggest that presence can function as a psychological bridge for integrating multiple modalities, such as visual, auditory, and tactile input. While previous studies show that multimodal integration can enhance the sense of presence, an important question remains: how can a cross-modal VR experience be designed to support or enhance presence, and in turn help users form stronger multi- and cross-modal associations? Therefore, effectively leveraging presence in VR could be pivotal for fostering richer cross-modal experiences, including those aimed at exploring or enhancing pitch-colour mappings.

Waterworth et al. (2012, (p. 190)) states that *"presence is not the degree of technological immersion, it is not the same thing as emotional engagement, it is not absorption or attention or action; but all of these and several other factors have a potential role in understanding the experience of presence in interaction"*. This view highlights presence as an emergent, multi-faceted phenomenon shaped by an interplay of sensory, cognitive, and affective processes. Meanwhile, Slater et al. (1998) demonstrate that active bodily engagement, such as physical movement and interaction with elements of the virtual environment, can significantly heighten presence, as shown in experimental tasks involving ecological inspection.

Designing for presence, therefore, requires more than technical optimisation. It involves balancing technical immersion, such as visual and auditory fidelity, tracking responsiveness, and minimal latency, with interaction quality, including locomotion, object manipulation, and user-driven exploration. It also includes affective and narrative elements that provide emotional depth and help users make sense of their actions and surroundings within a coherent virtual context. Presence is best cultivated by aligning hardware capabilities with user-driven, cognitively manageable interactions, and immersive (but not overwhelming) narrative frameworks. This multi-faceted approach supports a more sustainable and impactful sense of presence, particularly in applications that demand attentional focus, learning, or cross-modal integration.

In essence, presence acts as a key psychological bridge between the user and the virtual environment. It influences how deeply users invest in tasks, how emotionally engaged they become, and how consistently they perceive (or dismiss) the virtual world's logic, including cross-modal mappings. By combining immersive technologies with well-crafted interactions and thoughtful design of multisensory cues, VR can potentially foster powerful pitch-colour associations that feel natural and believable. However, designers must remain cautious about overloading users cognitively or emotionally, so that presence does not become overwhelming rather than beneficial.

By understanding both immersion (the system's objective qualities) and presence (the user's subjective perception), we can more effectively develop VR applications that adopts effective cross-modal design, particularly in contexts like pitch-colour associations. The next sections will examine how specific features, such as interactivity and spatial audio further shape the VR experience.

2.4.3 Interactions in VR

Interactivity is a key component of VR, shaping how users perceive and engage with a virtual environment. As previously discussed in Section 2.4.2, interactions can reinforce a sense of *"being there"*, complementing the technical aspects of immersion (Regenbrecht and Schubert, 2002). According to Bowman et al. (2001); LaViola et al. (2017), VR interactions can typically be broken down into four main categories: selection, manipulation, locomotion, and system controls. Mine (1995) highlights similar categories, with selection, manipulation, locomotion, and scaling; however, scaling can be viewed as a type of manipulation. Selection, Manipulation, and Locomotion are highlighted below as areas that can help foster cross-modal associations in VR environments, helping to influence a sense of presence and immersion for the user.

Selection

Selection involves choosing or activating objects within the VE: for instance, grabbing a virtual ball, pressing a menu button, or toggling an interface element (Mine, 1995). The most common selection methods include (Mine, 1995):

- Direct Interactor: Physically *"touching"* an object in VR (often via motion controllers or tracked hands).
- Raycaster: Using a laser pointer metaphor to select distant objects.
- Gestures: Hand-tracking technology identifies specific hand shapes (e.g., pinch) for user input.
- Gaze: Eye-tracking hardware detects where a user looks for a set duration, initiating selection.

Providing multimodal feedback (auditory, visual, haptic) is essential to confirm a successful action and reduce cognitive load, giving the user an immediate sense of agency and completion.

Manipulation

Once an object is selected, manipulation allows users to adjust it by rotating, scaling, or moving it (*Interaction Techniques in VR*, n.d.). Simple examples include throwing a ball (move or transform), opening a door (rotation), or zooming into a virtual map (scaling). Realistic physics and sensory feedback can deepen engagement by creating a convincing illusion that objects respond naturally to user input, thus reinforcing the sense of presence (Mangalam et al., 2024). However, more advanced manipulation methods (e.g., two-handed interactions, physics-based grips) may also increase cognitive and physical demands, so designers must balance immersion with usability, especially for longer or more complex tasks.

Locomotion

Locomotion governs how users move through the virtual space: via controller-based movement, teleportation, or walk-in-place, among others (Mine, 1995). Each

approach offers trade-offs regarding comfort, presence, and simulation sickness (Tomasetti and Turchet, 2024) (see Section 5.2.3 for further discussion). It has been found that participants preferred a choice of locomotion (e.g., switching between teleportation and free movement), as it allowed them to adapt movement style depending on the task context or personal comfort (Tomasetti and Turchet, 2024). Although flexible locomotion can enhance presence, studies have shown that dual tasking can increase cognitive demands (Kizony et al., 2010). Careful implementation is needed when designing the locomotion needs along with the other interaction types. Making the interaction as intuitive as possible is key to preventing negative effects, such as cognitive overload, while also enhancing presence, immersion, and engagement.

Interaction Design and Performance

Good 3D interaction design can improve task efficiency and user satisfaction compared to traditional mouse-and-keyboard setups (McMahan et al., 2006). Interestingly, McMahan et al. (2006) found that immersion level alone did not predict better performance; instead, interaction technique was key for tasks such as object manipulation. Concerning usability, some research suggests that dedicated controllers are more reliable and physically less demanding, whereas hand-tracking feels more intuitive for new users (Capece et al., 2023; Kojić et al., 2024). Designers should thus tailor interaction methods to application goals and user characteristics, balancing innovation and cognitive or physical demands.

Relevance to Cross-Modal Associations

Thoughtful interaction techniques has the potential to reinforce pitch-colour associations by linking user actions to immediate audio-visual feedback. For instance, the Chromaesthetica application demoed at the `Mobile and Ubiquitous Multimedia (MUM) conference`, highlighted user interactions with visual orbs that mapped to musical notes, adjusting the colour through controller movement until it aligns with the users visual perception of the sound O'Toole et al. (2024). These interactions aim to explore how pitch-colour associations are formed in a VR context. Through repeated interactions, users may come to internalise these pairings more strongly, particularly if they perceive the VE as a unified, believable space (further discussed in Section 2.5.1 on Multisensory Integration). Furthermore, manipulation (e.g., moving or resizing an orb) can allow the pitch or colour intensity to change in real time, enhancing multisensory integration and encouraging deeper engagement with the cross-modal mapping.

The next section discusses the importance of spatial audio in VEs. While spatial audio could fall under the previous discussions on immersion or presence, given audio's significant role in this research, it warrants a more focused section. Although audio's importance in VR has been overlooked in the past, it can be essential for enhancing immersion and presence in a virtual environment.

2.4.4 Spatial Audio

Spatial audio (or 3D audio) is another crucial dimension of immersive VR design. Defined as sound perceived to originate from specific points in three-dimensional

space, spatial audio significantly bolsters realism, presence, and multisensory coherence (Broderick, 2021). Although audio information can sometimes be processed more rapidly than visuals (Kailas and Tiwari, 2021), VR research and development traditionally prioritised visual fidelity over advanced audio rendering (Broderick et al., 2018).

From Stereo to 3D Audio

Conventional stereo setups provide only two channels (left and right) with no height or distance cues (*Mono vs Stereo Audio: Understanding the Differences*, n.d.). In contrast, spatial audio simulates how sound waves arrive at the listener's ears from multiple directions (Kailas and Tiwari, 2021). This is typically achieved through either binaural recordings, which capture audio with microphones placed in a dummy head to preserve natural localisation cues, or Head-Related Transfer Functions (HRTF) that use mathematical models to position sound dynamically, factoring in head shape, torso, and ear contours (Zhang et al., 2017; Kailas and Tiwari, 2021).

HRTFs can be individualised for maximum fidelity, but at a higher cost. Non-individualised datasets (e.g., `KEMAR`) are typically used in commercial VR for a balance of effectiveness and convenience (Padmanaban and Molner, 2016). A study by Parseihian and Katz (2012) showed participants can quickly adapt to non-individualised 3D audio for vertical sound location when given interactive training. Similarly, a study by Potter et al. (2022) comparing monaural audio, head-tracked binaural audio, and binaural audio with room acoustics found that increasing spatial audio fidelity significantly enhanced perceived immersion and audio-visual quality, with room-acoustic rendering producing the strongest effect. While spatial music has been shown to enhance immersion in VR, qualitative findings suggest that it may also introduce perceptual or attentional drawbacks (Mahlamäki, 2023). This can depend on context, indicating that conventional stereo music can remain an appropriate design choice in certain VR applications (Mahlamäki, 2023). Similarly a study by Brinkman Willem-Paul et al. (2015) indicated that stereo audio, while less spatially precise than 3D audio, can still deliver a heightened sense of presence in VR, supporting its use as a viable alternative in contexts where full spatialisation is unnecessary or impractical. This balance between 3D and stereo audio underscores the importance of considering how spatial audio can shape users' sense of presence and integration between auditory and visual modalities, particularly within immersive learning environments.

Linking Spatial Audio to Features of Multisensory Experiences

Recent research indicates that spatial audio can contribute to increased immersion, presence, cognitive performance and task performance in virtual environments. For example, Potter et al. (2022) demonstrated that head-tracked binaural audio, particularly when combined with high-resolution visual displays, significantly enhances users' self-reported sense of immersion. In terms of presence, spatial audio was found to have a medium to strong effect on participants' perceived sense of being in a virtual environment (Poeschl et al., 2013). With regard to performance within a VR setting, Gröhn et al. (2005) found that the inclusion of auditory and audio-visual feedback improves navigation efficiency within virtual environments. Spatial audio has also been shown to support cognitive processing by directing user attention to-

ward relevant events or objects within the environment, thereby reducing cognitive load (Huang et al., 2024). Collectively, these findings suggest that spatial audio plays a critical role in shaping how users perceive, attend to, and interact with virtual environments. Existing work has primarily examined these effects in terms of usability. However, spatial audio has the potential to anchor auditory information to specific locations in space, thus presenting opportunities to explore spatial audio in cross-modal associations. In particular, spatial audio may support the formation and reinforcement of auditory-visual associations, such as pitch-colour mappings, within virtual environments.

Relevance to Cross-Modal Associations

Integrating spatial audio with visual colour cues can potentially create richer cross-modal scenarios in VR. Studies indicate that audio and audio-visual stimuli increases users' sense of presence and immersion (Potter et al., 2022; Poeschl et al., 2013), and that adding spatial auditory cues to visual tasks enhances multisensory brain processing and behavioural performance (Alwashmi et al., 2024). With thoughtful design of VR environments, more immersive and cohesive cross-modal scenarios can be created using congruent spatial audio stimuli with visual features, such as colour. For example, suppose a note's pitch is not only linked to a corresponding hue but also emanates from a distinct position in 3D space. In that case, users may be more inclined to attribute *"realness"* to that pitch-colour connection. Coupled with interactive mechanics (e.g., selecting or manipulating orbs that emit both audio and colour changes), such an approach can harness presence, immersion, and multisensory integration to strengthen pitch-colour associations in an engaging, memorable way.

This section has explored VR's capabilities for engaging users through immersion, presence, interactivity, and spatial audio. Research into these features has demonstrated their benefits across various application areas, including task performance, cognitive load reduction, user understanding and engagement, and knowledge transfer. While these capabilities can support a wide range of goals, this thesis focuses on examining how they might help foster and support cross-modal associations between pitch and colour. Achieving this requires a balanced approach—one that draws on the strengths of each feature to inform the design of virtual environments capable of enhancing cross-modality and creating meaningful user experiences in specific learning contexts.

The next section turns to multisensory learning and engagement, highlighting key concepts such as multisensory integration, embodiment, affective environments, and cognitive load, and examining how they relate to the core research question of this thesis.

2.5 Multisensory Learning and Engagement

The real world is inherently multisensory; our senses and their interactions continuously inform our perception and understanding of our environment (Martin et al., 2022). Immersive VR systems offer the unique advantage of replicating multisensory real-world experiences, notably through visual, auditory, and haptic modalities (Martin et al., 2022). The previous section discussed key factors such as immersion,

presence, and interaction, emphasising how effectively designed multimodal feedback significantly enhances immersive experiences in VR. Recently, VR has become an important tool in training and educational contexts, offering realistic multisensory scenarios safely and cost-effectively (Bordegoni et al., 2023; Grabowski and Jach, 2021).

This section explores an example of how multisensory feedback impacts learning, memory, and cognitive processing in educational and musical settings. To understand how pitch-colour cross-modal associations can be supported in VR environments, this section draws on research from multisensory integration, embodied cognition, affective factors in multisensory learning, Cognitive Load Theory (CLT), and Cognitive Theory of Multimedia Learning (CTML). These theoretical perspectives provide insights into how associations between auditory and visual modalities are formed, offering useful perspectives on how pitch-colour pairings might be enhanced. Each contributes to understanding how users process, engage with, and form cross-modal associations in immersive VR settings.

2.5.1 Multisensory Integration and Cross-Modality

Multisensory learning involves engaging multiple sensory modalities simultaneously or sequentially to enhance the learning experience, reflecting the natural conditions of everyday perception (Alwashmi et al., 2024). Historically, educational practices often focused primarily on single-sensory modalities, typically visual or auditory alone (Shams and Seitz, 2008). However, research suggests that integrating stimuli from multiple senses significantly improves perceptual and cognitive performance both in the real world and in virtual environments (Alwashmi et al., 2024).

Multisensory integration is defined as the neural process by which the brain combines inputs from two or more sensory modalities, such as vision, hearing, and touch, into a coherent perceptual experience (Stein and Stanford, 2008). Effective multisensory integration enhances learning, particularly when sensory stimuli are temporally synchronised and semantically congruent (Shams and Seitz, 2008). Temporal synchrony refers to presenting multisensory inputs simultaneously or within a narrow temporal window, typically 100-200 milliseconds, facilitating their integration into a unified perception (Stevenson et al., 2012; van Wassenhove et al., 2007). Semantic and perceptual congruency further strengthens multisensory integration (Chen and Spence, 2010; Laurienti et al., 2004; Parise and Spence, 2009). Semantic congruency is achieved when stimuli from different senses logically match (e.g., a dog's image paired with barking), whereas perceptual congruency aligns stimuli based on perceived sensory qualities (e.g., low-pitched sounds matched with larger visual objects).

Cross-modal associations differ from multisensory integration as they involve subjective links rather than direct neural binding (Spence, 2011; Stein and Stanford, 2008). While these two processes naturally interact, VR environments offer unique opportunities to study and design how they work together. Synchronously presenting auditory pitch and visual colour within interactive tasks can leverage multisensory integration mechanisms, thus potentially reinforcing or creating robust cross-modal associations (Shams and Seitz, 2008; Parise and Spence, 2012). Evidence supports the notion that multisensory training surpasses single-sensory training in enhancing memory and learning effectiveness (Shams and Seitz, 2008).

2.5.2 Embodied Cognition in VR

Embodied cognition refers to the theory that cognitive processes are fundamentally grounded in bodily interactions with the environment, suggesting that our sensory and motor experiences directly influence cognitive functions such as perception, learning, and memory (Wilson, 2002). Central to embodied cognition is the idea that cognition is not solely brain-based, but also inherently linked to bodily actions, sensory experiences, and environmental contexts.

In immersive VR environments, embodied cognition becomes particularly significant due to the heightened potential for sensorimotor interactions. Unlike traditional screen-based technologies, VR allows users to physically interact with virtual objects and environments through naturalistic movements such as reaching, grasping, and manipulating objects directly, which closely replicate real-world interactions (Gallagher and Lindgren, 2015; Kilteni et al., 2012). These embodied interactions in VR not only enhance the realism and immersiveness of experiences but also support deeper cognitive engagement and more effective learning by activating embodied cognitive mechanisms.

Research has demonstrated that engaging bodily movements and interactions within learning tasks can significantly enhance memory retention, problem-solving ability, and conceptual understanding (Lindgren and Johnson-Glenberg, 2013). For instance, physically interacting with objects representing abstract concepts (such as pitch and colour associations) could facilitate a deeper understanding of these relationships by leveraging embodied cognitive processes. This can help learners internalise abstract or arbitrary associations more efficiently than through passive observation or non-interactive presentations.

Additionally, research into embodied cognition suggests that bodily actions and gestures in VR can reduce cognitive load by offloading cognitive demands onto sensorimotor systems. This effect occurs because physical interactions provide additional sensory cues, reinforcing cognitive associations through multimodal feedback, thus facilitating the formation of cognitive schemas (Pouw et al., 2014). For example, imagine a virtual piano where each key lights up with a specific colour as it is played, with higher-pitched notes appearing in lighter colours and lower-pitched notes in darker tones. By engaging with this kind of interactive, multisensory setup, users begin to form intuitive links between pitch and colour through embodied experience.

Embodied interactions also strongly intersect with multisensory integration. By combining tactile, visual, and auditory feedback within a single coherent action, VR systems can reinforce multisensory congruency and integration, thereby enhancing the perceived realism and saliency of the learning environment (Gallagher and Lindgren, 2015). Therefore, embodied interactions can actively promote stronger multisensory and cross-modal associations, which are essential for investigating pitch-colour mappings in VR.

In summary, embodied cognition provides a way for understanding how bodily interactions in VR can significantly enhance cognitive processes, reduce cognitive load, and support the formation and reinforcement of cross-modal associations. Carefully designed embodied interactions that leverage multisensory cues hold substantial promise for improving pitch-colour learning tasks in VR environments, directly addressing this thesis's research goals and questions.

2.5.3 Affective Factors in Multisensory Learning

Affect and emotion can significantly influence core cognitive processes such as attention, memory, and learning. According to Pekrun and Stephens (2012), emotional engagement can enhance or hinder learning by affecting cognitive resource allocation, motivation, and the use of learning strategies. Positive emotions, in particular, are associated with heightened attention and improved memory consolidation, both of which are crucial for effective learning in complex environments, such as VR.

In multisensory learning contexts, congruent sensory stimuli can enhance multisensory integration by reinforcing associative links and improving recall (Shams and Seitz, 2008). Separately, research in educational psychology shows that emotional engagement enhances attention, memory encoding, and learning outcomes (Pekrun and Stephens, 2012; Pessoa, 2008). Thus, emotional engagement is not merely an affective experience but also a cognitive tool that strengthens learning through deeper processing.

VR environments are especially potent for eliciting emotional responses due to their immersive and interactive nature. Riva et al. (2007) demonstrated that VR could be used to induce strong emotional states, which in turn affected presence and cognitive engagement. Similarly, Pallavicini et al. (2019) found that body-involving VR experiences elevated positive emotional responses compared to non-immersive conditions.

Building on these findings, O'Toole et al. (2021) (P1) explored the emotional associations triggered by single musical notes. This study demonstrated that even minimal auditory stimuli could evoke consistent emotional responses (e.g., lower pitches linked with sadness, higher pitches with joy), suggesting that simple, well-chosen auditory stimuli can act as powerful emotional drivers. Importantly, this supports the idea that pitch stimuli used in VR tasks, especially when synchronised with colour changes, can evoke emotional responses that enhance multisensory engagement and memory formation.

Therefore, in the context of this thesis, carefully designing VR tasks with emotionally resonant cues, such as matching pleasant musical pitches with appealing colour shifts, could increase user engagement and support more robust cross-modal association formation (see Chapter 6). Emotionally enriched multisensory learning environments may therefore represent a key mechanism by which VR can facilitate enhanced pitch-colour connections.

2.5.4 Cognitive Load Theory (CLT)

Managing cognitive load is crucial for effective learning, particularly in immersive VR environments. CLT, initially proposed by Sweller (2020), describes how instructional design impacts the efficiency of working memory during learning processes. CLT categorises cognitive load into intrinsic (task complexity), extraneous (load due to suboptimal instructional design), and germane (productive mental effort dedicated to learning) types. Effective instructional design aims to minimise extraneous load, manage intrinsic load, and maximise germane load to facilitate schema formation and deeper understanding (Sweller, 2011).

CLT research highlights several cognitive load effects that instructional designs can exploit to improve learning outcomes. Specifically relevant to VR environments are the split-attention, modality, and transient information effects. The split-

attention effect occurs when learners must divide their attention across spatially or temporally separated information sources, thereby increasing extraneous load and decreasing learning efficiency (Sweller, 2020). Integrating pitch and colour cues spatially and temporally within VR systems helps minimise this effect.

The modality effect occurs when related information that would normally cause split attention can instead be processed more effectively by distributing it across visual and auditory channels (Sweller, 2020). By presenting one source of information visually (e.g., diagrams) and the other via an auditory source (e.g., spoken text), working memory capacity is increased, thereby enhancing learning (Sweller, 2020). In VR-based pitch-colour learning tasks, combining auditory instructions or pitch cues with visual colour changes exploits dual-channel cognitive capacities, reducing cognitive overload.

The transient information effect describes increased cognitive load that occurs when instructional content is fleeting or transient, making it challenging for learners to process adequately (Sweller, 2020). VR environments can address this by incorporating persistent visual or replayable auditory cues, allowing learners sufficient processing time.

Following CLT guidelines can thus significantly enhance the design of VR-based educational tools, reducing unnecessary cognitive load and improving learners' ability to form and retain cross-modal associations effectively.

2.5.5 Cognitive Theory of Multimedia Learning (CTML)

CTML, proposed by Mayer (2005), complements CLT by focusing specifically on multimedia environments. CTML argues that individuals learn more effectively when combining words (spoken or written) with visual content, provided that instructional materials respect cognitive processing limitations (Mayer, 2023).

CTML is founded on three core assumptions: dual-channel processing (auditory and visual), limited working memory capacity, and active cognitive processing (Mayer, 2005). Mayer (2023) outlines several instructional design principles to manage cognitive load in multimedia contexts, such as the modality principle (presenting information via dual channels), redundancy principle (avoiding identical auditory and textual content), segmentation (breaking information into smaller, manageable units), pre-training (providing foundational knowledge in advance), and signalling (highlighting key information to guide learners).

Although CTML's principles are not explicitly tailored for cross-modal pitch-colour associations, they directly inform the instructional design strategies used in VR environments explored in this thesis. Adopting CTML principles, such as segmentation, pre-training, and signalling, can significantly reduce cognitive overload, thereby enhancing the learner's ability to effectively form and retain cross-modal associations.

2.5.6 VR as a Tool for Cross-Modal Learning

VR's inherent capability to integrate multisensory stimuli, manage cognitive load effectively, and provide embodied and emotionally engaging experiences positions it uniquely as a powerful medium for cross-modal learning. Prior studies highlight both the potential and challenges of VR in educational contexts. Studies by Makran-

sky et al. (2019) and Schrader and Bastiaens (2012) indicate that poorly designed immersive experiences can increase cognitive load, thereby negatively impacting learning outcomes. Conversely, research suggests that well-designed VR environments can significantly improve multisensory engagement, cognitive efficiency, and learning performance (Martin et al., 2022; Luong et al., 2021; Pallavicini and Pepe, 2019). With regard to learning, performance, engagement, and cognitive load, careful consideration is required in the design of virtual learning environments that have the potential to strengthen cross-modal connections between the auditory and visual senses.

Studies have shown that auditory-visual cross-modal associations are present in VR settings (Malpica et al., 2020; Larsson et al., 2002). However, they only examined general associations between the two senses and did not explore how features inherent to VR might impact, influence, or strengthen the connection between modalities. Kim et al. (2023) demonstrated that auditory-visual cross-modal associations affect both how quickly and how strongly users experience vection—the illusion of self-movement—in VR. This highlights how cross-modal connections can enhance user experience within virtual environments, in this case by potentially reducing motion sickness. These findings highlight a dual perspective on cross-modal associations in VR. Beyond examining how VR can strengthen pitch-colour associations, cross-modal associations can be leveraged to enhance overall user experience within virtual environments.

As previously highlighted in the other subsections in Section 2.5, multisensory integration, embodiment, affective environments and cognitive load can play important roles in the design of VR environments, balancing immersion, presence, engagement and learning ability. A recent study proposed exploring reduced realities (RR) that harness cross-modal design to more effectively balance the positive and negative aspects of VR. The advantage of this approach is that it creates the desired amount of presence, while avoiding the stress that can come with a highly immersive and cognitively taxing environment (Grimshaw-Aagaard and Walther-Hansen, 2024). A study by Williams et al. (2015) highlights that the best learning happens when all our senses work together in a congruent and meaningful way. This kind of integrated sensory experience can help build a better understanding that isn't limited to the auditory or visual senses. VR as experienced through a HMD is better placed than traditional 2D technologies, such as standard computer setups, to support multi- and cross-modal relationships and to explore how virtual environments can support learning. Therefore, designing VR environments to facilitate cross-modal learning should involve carefully balancing immersive elements, interactivity, cognitive load management, and multisensory integration principles. Effective VR designs that leverage these components can create intuitive, memorable, and cognitively manageable learning experiences.

In summary, multisensory integration, embodied cognition, affective factors, cognitive load theory, and cognitive theory of multimedia learning provide critical guidance for designing effective VR learning experiences. Each of the aforementioned theories highlights different but complementary factors. They show how the senses can be combined to form coherent perceptions, how information can be structured across multiple modalities to reduce cognitive overload, how interactions can support cognitive processes, and how multimodal content can enhance understanding. These insights are particularly relevant when exploring pitch-colour cross-modal associa-

tions, where auditory and visual stimuli must be presented in a coherent, engaging, and cognitively manageable way. For instance, synchronising pitch and colour cues through interactive tasks can enhance perceptual binding (via multisensory integration), reduce unnecessary mental effort (via CLT), and support learning through meaningful bodily engagement (via embodied cognition). By leveraging VR's unique capabilities, designers can effectively foster cross-modal associations, such as pitch-colour mappings, offering immersive and enriched educational environments.

2.6 Cross-Modal Research in XR Environments

While Section 2.2 has provided an overview of cross-modal associations in real-world contexts, this section focuses explicitly on the state of cross-modal research within XR environments. Although previous sections (Section 2.4) have touched upon the potential benefits of cross-modal interactions in VR, this section focuses on existing research that explicitly investigates cross-modal perception in immersive settings.

The primary strength of VR lies in its multimodal capabilities, delivering immersive experiences that engage multiple senses—primarily visual, auditory, and haptic modalities (Martin et al., 2022). The recent rise in popularity and accessibility of VR technologies has generated extensive interdisciplinary research, typically categorised into two broad streams: VR as a methodological tool to investigate specific phenomena, and VR as the subject of inquiry itself. While multimodal integration—the simultaneous engagement of multiple senses—has been extensively studied (Martin et al., 2022; Spence, 2007; Stein and Stanford, 2008), explicit research into cross-modal perception, where one sensory modality influences another, remains relatively scarce in immersive VR environments. The following subsections explore current knowledge and highlight gaps within cross-modal research specific to XR settings.

2.6.1 VR as a Platform for Multisensory & Cross-Modal Experience

Earlier discussions have illustrated VR's capabilities for delivering immersive, realistic experiences that enhance presence, memory, and learning outcomes through multisensory integration (Sections 2.4 and 2.5). Building on this, recent research has begun exploring the unique advantages of VR's multimodal capabilities in various educational and training scenarios. For instance, Bordegoni et al. (2023) demonstrated positive outcomes in XR-based industrial machinery training, language learning, and music education, highlighting the effectiveness of multimodal interactions for skill acquisition. Other studies, such as those conducted by Poguntke et al. (2019) and Gradl et al. (2019), have applied VR to cognitive research using the Stroop test, an established psychological paradigm involving congruent (e.g., the word "Green" printed in green) and incongruent colour-word pairs. Gradl et al. (2019) found that VR versions of the Stroop test provoked higher physiological stress responses compared to desktop versions, underlining VR's effectiveness in creating controlled stress conditions. Mevlevioğlu et al. (2023) also found that VR has the potential to measure different anxiety states when conducting an emotional Stroop word colour task. Conversely, Poguntke et al. (2019) revealed that VR can also provide reduced

stress conditions compared to traditional desktop environments, particularly when no additional stressors (like head movements) are included. Together, these studies demonstrate VR's versatility, underscoring its potential to balance emotional states and cognitive conditions, making it highly beneficial for stress-related research and applications that require careful sensory management.

Regarding auditory-visual associations in VR, research targeting music education and creativity has provided valuable insights. A study by Ppali et al. (2022) investigated the possibility of employing VR to support musicians in rehearsal, improvisation, and creative practices, suggesting potential benefits of multisensory integration in musical contexts. Another study proposed a VR-based system for teaching music composition through visual representations of harmony and chord progressions, further emphasising the potential of VR for enriching musical education (Timoney et al., 2018). However, despite acknowledging VR's multimodal qualities, these studies have not deeply explored the explicit relationships between auditory and visual stimuli (e.g., pitch-colour associations), leaving important questions unanswered about how cross-modal correspondences could be enhanced in VEs and how they can benefit learning and creative outcomes. With studies examining the effects of affective responses (Riva et al., 2007; Pallavicini et al., 2019), embodiment (Lindgren and Johnson-Glenberg, 2013), and cognitive load (Rettinger et al., 2024; Pouw et al., 2014) on users' perception of virtual environments, it is evident that there is more to consider beyond just direct pitch-colour associations alone. Incorporating these factors into virtual environment design may support more effective multisensory integration. Addressing these gaps represents a key aim of this thesis, which specifically investigates how VR can facilitate stronger and more meaningful pitch-colour cross-modal associations.

2.6.2 Evidence of Cross-Modal Correspondences in XR Settings

Although multimodal interactions have received significant attention, explicit studies of cross-modal correspondences in VR—particularly auditory-visual mappings—remain limited. However, recent work has begun exploring how XR settings might deliberately enhance cross-modal perception. For instance, Malpica et al. (2020) demonstrated that auditory-visual cross-modal effects, well-established on conventional displays, also occur in VR, though slightly weaker in strength. Notably, their findings suggest that well-designed auditory cues can compensate for lower visual fidelity, providing enhanced perceptual realism without substantial graphical rendering requirements (Malpica et al., 2020).

Larsson et al. (2002) confirmed the presence of auditory-visual interactions in virtual environments, although they noted that much of the research in this area remains exploratory. They called for more targeted perceptual studies of specific cross-modal effects (e.g., how sound shapes spatial perception) and recommended using affective and physiological measures alongside subjective reports. These points highlight the importance of considering simulation fidelity and presence when investigating cross-modal correspondences, which is relevant for exploring pitch-colour associations.

Further supporting these findings, Grimshaw-Aagaard and Walther-Hansen (2024) introduced an *"audio-first"* design philosophy, demonstrating that auditory

stimuli could effectively guide user attention, support emotional engagement, and prevent sensory overload. This approach emphasises the practical value of auditory-visual cross-modal design, particularly in scenarios prioritising focus and cognitive clarity, such as educational and therapeutic VR applications.

Research into perceptual phenomena like vection—the illusion of self-motion—also provides valuable evidence. Kim et al. (2023) found that auditory-visual congruency significantly influenced the sensation of self-motion in VR, with both congruent and incongruent auditory stimuli strengthening the visual perception of movement. This demonstrates how cross-modal cues can strongly shape perception, implying that thoughtful pitch-colour mappings could similarly enhance user focus, immersion, and learning in VR.

Artistic and narrative applications have further explored the potential of cross-modal and synaesthetic experiences in VR. Haverkamp (2022) explored challenges and opportunities in digitally representing synaesthetic experiences, arguing that effective synaesthetic art must transcend literal translation to achieve broader artistic impact. Similarly, Rudenko et al. (2022); Rudenko and Cabral (2019) examined how cross-modal insights from synaesthetic perception can enhance digital applications in AR and VR, ranging from educational and therapeutic interventions to narrative visualisation of classical music. Weinel (2021) further demonstrated practical strategies through interactive VR *"sound toys"*, allowing users to dynamically engage with and manipulate cross-modal stimuli in real-time, thereby significantly enhancing sensory and creative immersion.

Taken together, these studies affirm the viability and value of cross-modal design within XR environments. They demonstrate how carefully constructed auditory-visual mappings can enhance perceptual realism, reduce cognitive load, and improve emotional and creative engagement. This body of work provides essential context for the current thesis, which investigates how VR-supported pitch-colour associations can be leveraged effectively to enhance multisensory learning and interaction.

2.6.3 Challenges & Considerations for Cross-Modal VR

Despite recent advances, designing effective cross-modal VR experiences remains challenging, particularly in relation to pitch-colour associations. Key considerations include ensuring temporal synchrony and perceptual congruency between auditory and visual inputs, as discrepancies can impair user engagement and perceived realism (Larsson et al., 2002). Furthermore, individual differences in sensory perception mean that users may intuitively map pitch to colour differently, necessitating approaches that allow personalisation or adaptability within VR environments.

Cognitive load represents another critical concern; overly complex multisensory environments can inadvertently overload users, reducing task clarity and educational effectiveness. Thus, balancing multimodal richness with cognitive demands is essential. Additionally, technical limitations in VR hardware, such as latency, audio fidelity, and rendering quality, can significantly influence cross-modal perception, necessitating careful consideration during the design process (Malpica et al., 2020). Lastly, designers must face conceptual tensions between fixed mappings (e.g., precise pitch-colour pairings) and flexible, interpretative associations, which may better support creative exploration and deeper engagement (Haverkamp, 2022).

Addressing these challenges is important to developing effective cross-modal VR

experiences that are immersive yet cognitively manageable, realistic yet creatively flexible, and universally appealing yet individually adaptive. These considerations can ultimately inform VR applications that meaningfully support cross-modal learning, interaction, and sensory exploration.

In addition to immediate design implications, acknowledging these challenges opens avenues for future research and practical application. Exploring personalised adaptive systems, sensory processing for diverse user groups, and innovative artistic and educational cross-modal applications represents valuable future directions. Through continued investigation, cross-modal VR design can contribute meaningfully to exploring improved learning, creativity, and sensory engagement.

2.7 Conclusion

When seeking to answer the question, *"How can the cross-modal associations of pitch and colour be enhanced through virtual reality?"*, several areas of exploration need to be considered. This chapter has presented a broad overview of the literature relevant to that central research question. While Chapters 4, 5, and 6 each contain specific background sections relating to their respective studies, this chapter has taken a broader view, connecting key ideas and themes across different fields to support the overall aim of the thesis.

The key takeaways from this chapter concern three main areas: how cross-modal associations and synaesthesia are understood; which features of VR can help users form meaningful sensory connections; and how multisensory learning techniques can support the integration and reinforcement of these associations in learning contexts. While pitch-colour associations appear to be widely shared across the population—including among synaesthetes—factors such as emotion or affect have been shown to influence how these associations are formed. Given the similarities between pitch-colour associations and chromaesthesia, this thesis adopts the position that the two may exist along a continuum, with differences in automaticity helping to distinguish them.

To explore pitch-colour associations using VR, this chapter examined the capabilities of the technology to identify which features and how they might inform the design of environments that support and enhance these associations. With limited research on pitch-colour cross-modal associations in VR, the viability of the technology for these perceptual connections must first be explored. This is highlighted in Section 1.3, with the research question that forms the focus of the first study in Chapter 4. Once VR can be confirmed as a viable pitch-colour testing tool, a more in-depth look at immersion, presence, interaction, and spatial audio is needed. The benefits and limitations of each feature have been considered. These elements are closely connected: immersion can influence presence, interaction can affect immersion, and so on. As such, designing a VR experience to support pitch-colour associations requires a balanced approach that considers how each element contributes to the overall experience, both individually and in relation to the others. This research is highlighted by RQ2, presented in Section 1.3, and investigated in the study in Chapter 5. Exploring the role of immersive features in forming pitch-colour cross-modal associations in VR is complex, involving a range of benefits and trade-offs depending on the specific features used and how they interact. Research has shown that, when used well, these features can improve learning outcomes, en-

hance performance, and reduce cognitive load. However, how best to apply these features specifically for cross-modal design is still not well understood, making it one of the goals of this thesis.

Another important focus of this chapter was on multisensory learning. In addition to understanding pitch-colour associations and the capabilities of VR, it was important to explore how learning theories might support the design of effective virtual learning environments. Several theories were discussed in this context, including multisensory integration, embodied cognition, affective engagement, CLT, and the CTML. Each of these theories offers useful insights: multisensory integration helps explain how different sensory stimuli can be combined; embodied cognition emphasises the importance of physical interaction; emotional factors show how engagement can support memory and association; and CLT and CTML provide practical strategies for managing cognitive load in immersive settings. Together, these theories provide a strong foundation for the design of VR tasks that aim to enhance cross-modal associations. The focus on multisensory learning, and the cognitive theories used to inform this design, underpin RQ3 (see Section 1.3), which is explored in Chapter 6 through the use of different pitch-colour mappings to assess learning outcomes in music exercises. This offers a practical use case for the central question of the thesis—how pitch-colour associations can be enhanced—within a domain closely connected to prior research in cross-modal associations and synaesthesia, music.

Although some previous research has explored cross-modality in VR, very few studies have focused specifically on pitch-colour associations. This presents a valuable opportunity to contribute new insights into how VR environments can be designed to support this type of sensory learning. The existing literature shows that cross-modal approaches in VR can enhance immersion, influence perception, and increase user engagement in meaningful ways. Building on these insights, the following chapters seek to address this gap by investigating how different design features in VR may affect pitch-colour cross-modal associations. To do so, the next chapter outlines the methodological framework used in this thesis, detailing the design, rationale, and approach taken across the three empirical studies.

Chapter 3

Methodology

3.1 Introduction

This chapter outlines the overarching methodology employed across the published studies that are presented in this thesis. Each study includes its own focused methodological detail—presented in Chapter 4 (Section 4.3); Chapter 5 (Section 5.3); and Chapter 6 (Section 6.3). The purpose here is to articulate the rationale behind the methodological approach taken across the thesis, in service of the central research question: *How can the cross-modal associations of pitch and colour be enhanced through virtual reality?*

As introduced in Chapter 1.3, each study contributes to addressing this overarching question by exploring a specific sub-question. This chapter discusses how the design of each study supports the broader research aims. It also explains the rationale for using a mixed-methods approach, in which quantitative scales and closed-ended questionnaire items were used for statistical analysis, while open-ended responses were analysed qualitatively. Alongside this, it provides a comparative overview of the methods used across the studies. The chapter begins by outlining the epistemological position of the researcher and the philosophical framing of the research. Epistemological positioning can be understood as the stance the researcher takes on the nature of knowledge, including how research questions are framed, how information is gathered, and how that information is interpreted (Moon and Blackman, 2017).

3.2 Epistemological Positioning

This thesis adopts a **pragmatist epistemological stance**, which prioritises practical inquiry and methodological flexibility over rigid adherence to any single philosophical paradigm. Pragmatism places the research question at the centre of the inquiry and supports the use of both qualitative and quantitative approaches as necessary to generate sound, actionable knowledge (Feilzer, 2010; Morgan, 2007). This direction is appropriate for interdisciplinary, design-based research such as this investigation into cross-modal pitch-colour associations in VR, which draws upon perspectives from psychology, human-computer interaction, music, and education.

Pragmatism does not reject epistemological traditions entirely; instead, it recognises that different research contexts may call for different ways of knowing (Hall, 2013). To situate this approach within the broader epistemological landscape, it is

helpful to briefly acknowledge three core positions: objectivism, subjectivism, and constructionism (Moon and Blackman, 2017). Objectivism posits that knowledge exists independently of human perception and can be discovered through objective observation. Subjectivism, in contrast, views meaning as entirely internal, arising from the individual's interpretation. Constructionism offers a middle ground, suggesting that knowledge is constructed through the interaction between individuals and the world, shaped by experience, culture, and context.

Constructionism elements inform this research's design, particularly in relation to how participants engage with VR elements to create pitch-colour mappings. Papert and Harel (1991) situates constructionism as a theory that emphasises that individuals learn most effectively when actively involved in creating personally meaningful artefacts. In Study Three (see Chapter 6), this principle is reflected in the experimental comparison between pre-defined and personalised mappings, where participants co-construct associations through their engagement with immersive, interactive learning environments. Such tasks support a constructionist view of learning as situated, experiential, and user-driven.

The pragmatist foundation of this research promotes methodological pluralism by employing statistical testing to measure performance outcomes and gathering open-ended user feedback to capture the subjective contextual experience. By embracing both approaches, the research provides a richer understanding of how pitch-colour associations are perceived, developed, and potentially strengthened through immersive design. Rather than privileging one type of knowledge, the thesis values practical insight and methodological appropriateness in answering the research question.

3.3 Research Design and Rationale

The research design of this thesis is structured around three interrelated studies, each contributing to the central research question posed in Section 3.1: *"How can the cross-modal associations of pitch and colour be enhanced through virtual reality?"* To address this overarching question, the studies were conducted in a progressive sequence, each building on the insights of the previous one.

The first study was a pilot designed to compare how people performed pitch-colour associations in VR and desktop environments. The second study explored how specific immersive features of VR influence engagement with pitch-colour associations. The third and final study implemented these design insights into a VR music learning environment, comparing personalised and pre-defined pitch-colour mappings regarding their educational impact.

3.3.1 Study One:

The research question addressed in the first study was: *"Is VR a viable platform for conducting pitch-colour association tests compared to desktop environments?"* Before investigating design-specific aspects of VR, it was necessary to establish whether the medium (a VR HMD) influenced performance on pitch-colour tasks. By comparing participants' scores on a web-based pitch-colour association test delivered in both desktop and VR HMD formats, this study provided a foundation for using VR in perceptual testing. This study highlighted several aspects of VR that helped

inform the design of Study Two, with the features of focus, immersion, and audio design at the forefront. The detailed methodology for this study is found in Section 4.3.

3.3.2 Study Two:

Building on the results of Study One, the second study explored the question: *"What role do immersive features play in supporting pitch-colour cross-modal associations in VR learning environments?"* The goal was to identify how environmental design, locomotion, interaction, auditory and visual presentation influence user immersion and task focus. This usability study informed the design of the final learning application by identifying VR features that enhance user engagement with cross-modal stimuli (detailed methodology found in Section 5.3).

3.3.3 Study Three:

The third study investigated the question: *"How do personalised and pre-defined pitch-colour cross-modal mappings compare in supporting learning outcomes during basic music theory exercises?"* To reiterate, personalised mappings refer to colours chosen by participants that they felt matched each note, whereas pre-defined mappings used specific colours already assigned to each note. This study focused on applying cross-modal design principles to a learning context, with a detailed methodology presented in Section 6.3. By comparing participant performance across both mapping types, the study aimed to evaluate whether personalised associations provide a more effective learning experience than structured, pre-defined mappings.

Together, these three studies provide a clear roadmap for addressing the central research question. Study One evaluates the basic perceptual viability of VR for pitch-colour cross-modal association tasks. Study Two investigates the immersive design factors that influence cross-modal engagement. Study Three tests the practical application of these associations in a music education context. This sequential design enables a comprehensive exploration of how virtual reality can support and enhance pitch-colour associations, from perceptual testing to interactive learning.

3.4 Mixed Methods Approach

A convergent mixed methods design was used in each study, combining quantitative and qualitative data to address the research questions (Creswell, 2022). In line with a pragmatic epistemological stance, the research prioritised the practical understanding of the phenomenon over allegiance to any single methodological paradigm (Feilzer, 2010), while drawing on constructionist insights where knowledge is constructed through creating, testing, and modifying physical or digital artifacts (Papert and Harel, 1991). Both the quantitative and qualitative data were analysed independently, and the findings were then interpreted together to provide a more comprehensive understanding of the results and to address the study's research questions more effectively.

While the first study, detailed in Chapter 4, was originally published as a short quantitative-only paper (P1), the chapter presents a fuller account of the study using the complete dataset. Chapter 4 provides further quantitative analysis of

the relationship between pitch and brightness, illustrated in Figure 4.3. Additionally, it includes qualitative insights gathered from post-study questionnaires that could not be included in the original publication because of space constraints. The studies presented in Chapters 5 and 6 were originally designed and published as mixed-methods investigations, incorporating both qualitative and quantitative components. The thesis chapters present these studies as they were (or intend to be) published, without changes to their methodology or analysis.

Given the interdisciplinary nature of this research—spanning HCI, Psychology, Music, and Education—investigating a perceptual phenomenon such as cross-modal associations in VR requires both objective measurements and subjective insights. While synaesthetic associations are often highly individualised, cross-modal associations tend to show common patterns across people. For example, higher pitches are frequently linked with brighter colours (Ward et al., 2006). However, individual differences still exist, such as variation in the specific hue chosen or the emotional meaning attached to the sound (Palmer et al., 2013).

If cross-modal perception is viewed as a weaker or learned form of synaesthesia, then it is expected that participants will experience these associations differently. Therefore, capturing both measurable performance outcomes and participants' subjective experiences is essential for developing a comprehensive understanding of how such associations are formed and strengthened in VR. The following section highlights the quantitative measurements and the statistical tests used to interpret the data across the three studies.

3.4.1 Quantitative Methodologies

Various quantitative measures were used across the three studies to evaluate participant performance, usability, and cross-modal engagement in different conditions. The selection of appropriate statistical tests was guided by established quantitative methodological principles (Greene and D'Oliveira, 2011). These measures were chosen to align with each study's aims and design, and included both objective task-based performance metrics and subjective rating scales. Table 3.1 summarises the measures and statistical tests employed.

In Study One (see Section 4.3), participant performance on pitch-colour association tasks was assessed through consistency scores derived from a standardised test conducted in both desktop and VR conditions. In addition, Red, Green, Blue Colour Model (RGB) colour values were collected to evaluate pitch-brightness associations. A paired samples t-test was used to compare consistency scores across the two conditions, while Fisher's exact test was applied to categorical outcomes due to the small sample size. To explore relationships between pitch and brightness values, Spearman's correlation analyses were conducted to account for potential linear and non-linear trends.

Study Two (see Section 5.3) focused on comparing subjective user experience across two VR environments. Participants completed several validated questionnaires, including the `System Usability Scale (SUS)`, `iGroup Presence Questionnaire (IPQ)`, `Cognitive Load Measurement (CLM)`, and `Simulator Sickness Questionnaire (SSQ)`. These measures provided a multi-faceted view of how immersive features influenced user engagement, comfort, and task focus. Given the ordinal nature of the Likert-scale data, the Wilcoxon Signed-Rank Test

was used for most comparisons. A repeated measures ANOVA was applied to the SSQ data to explore trends in mean scores across conditions.

In Study Three (see Section 6.3), performance in the music exercises was assessed through error scores and task completion times in a between-subjects design. Participants were divided into two groups, one using pre-defined pitch-colour mappings, and the other using personalised mappings. The Mann-Whitney U test was used to compare the two groups' performance (error scores and time taken). In addition, a repeated measures ANOVA was used to examine trends in performance across lessons and between groups, specifically for error scores and task completion time. These analyses were used to explore broader patterns across the learning sequence.

This study-specific approach to quantitative analysis ensured that the chosen statistical methods were aligned with data type, distribution, and research design, supporting valid and reliable comparisons across experimental conditions.

From a qualitative perspective, participants' experiences were explored through open-ended feedback questions included in the post-study surveys. This allowed for a deeper investigation into how users perceived and engaged with the pitch-colour association tasks. A template analysis was applied to analyse this feedback, enabling the organisation of user reflections into structured thematic categories. The following section provides a detailed justification for selecting template analysis over other qualitative methods.

3.4.2 Qualitative Methodology

Template analysis is one of many forms of thematic analysis that can be conducted in qualitative research. It involves generating and refining codes in an iterative process, which are then organised into categories to develop overarching themes (the specific procedures are described in more detail for each study in Chapters 4, 5, and 6) Brooks et al. (2015). It was adopted to help gain meaning from the qualitative data gathered from the post-study questionnaires completed by participants in all three studies.

Brooks and King (2012) states that a *"template analysis can be used to pragmatically address real world issues in a wide variety of settings"*. A template analysis is a flexible tool which allows for themes to be defined *"a priori"*, i.e., in advance. While the use of *"a priori"* themes is not required, much research starts with assumptions regarding the focus of the study area (Brooks and King, 2012). With the pragmatist epistemological standpoint (highlighted in Section 3.2), the use of the template analysis for the qualitative aspects of the studies is beneficial when working in real or, in this case, virtual settings where perceptual tasks and users' experience are involved.

Given the pragmatist stance of this research, several qualitative thematic analysis approaches were considered. While other forms of thematic analysis, such as the reflexive thematic analysis, offer a flexible framework for identifying patterns in qualitative data, they typically do not use pre-defined codes (Braun and Clarke, 2019). Template analysis, on the other hand, was chosen for its ability to incorporate *"a priori"* codes while still allowing flexibility to adapt and refine themes as the analysis progresses. Unlike grounded theory, which is primarily concerned with developing new theory from data (Corbin and Strauss, 2014), template analysis was better suited to this study's design-focused and evaluative nature.

Other qualitative methods were explored including Interpretative Phenomenological Analysis (IPA). However, this approach generally used to explore how people deal with significant life experiences. It is usually used in studies regarding health, and is rich in capturing individual lived experiences, usually dealing with a smaller sample size (Smith et al., 2021; Alase, 2017). This approach was less applicable for the studies presented in this thesis, given the sample sizes, captured data and comparative aims of the research.

Template analysis enabled the use of *"a priori"* codes based on research objectives (e.g., immersion, mapping clarity, task engagement), while still allowing for new themes to emerge (Brooks and King, 2012). In each of the studies presented in Chapters 4, 5, and 6, the initial ideas based on the research goals served as the foundation for the *"a priori"* codes. Through a process of data exploration and iterative refinement of themes and codes, the final themes presented in this thesis were identified (see Figures in Appendices A.6, B.9, and C.6). This iterative process and the flexibility to adapt the coding made it particularly appropriate for analysing open-ended feedback across multiple conditions in a structured yet adaptable way. All participant feedback from the post-study questionnaires was included verbatim in the template analysis across the three studies in this thesis.

3.5 Overview of Methods Used in Each Study

Table 3.1 summarises the key methodological features of this thesis's three studies. Studies One and Two adopted a within-subjects A/B testing design, allowing participants to experience both conditions (e.g., desktop vs VR, or the forest vs space virtual environments). This approach reduced individual variability and was appropriate for perceptual and usability-focused investigations. In contrast, Study Three employed a between-subjects design to compare personalised and pre-defined pitch-colour cross-modal mappings in a learning task. Participants completed a sequence of eight progressively challenging music theory lessons using only one mapping type. A within-subjects design was avoided to prevent confusing the learning progression with exposure to multiple mapping systems, which could have introduced cognitive interference, biased learning outcomes, or diluted the comparative value of each condition.

Details on sample selection, participant demographics, and data collection procedures are presented within each of the three empirical study chapters (Chapters 4, 5, and 6), where methodological decisions are discussed in relation to the specific aims and context of each study.

Across the studies, a combination of structured quantitative measures and qualitative feedback was employed, particularly in Studies Two and Three, where participant experience and engagement played a central role. Integrating both methodological strands provided a more nuanced understanding of how virtual environments influence cross-modal associations and music learning. Together, these methodological choices reflect a deliberate, staged approach to investigating the thesis's central question across perceptual, experiential, and applied domains.

Table 3.1: Summary of Studies on Auditory-Visual Testing in VR Environments

Study Title	**Paper / Chapter**	**Design Type**	**Users**	**Quantitative Analysis**	**Qualitative Methods**	**Qualitative Analysis**	**Setting / Environment**	**Instruments / Measures**	**Main Aim**
Auditory-Visual Test	Paper 1 / Chapter 4	Within-subjects A/B test	20	Paired T-Test, Fisher's Exact Test, Pearson's and Spearman's Test	Open-ended feedback	Text Analysis and Template Analysis	Desktop vs VR HMD (web-based pitch-colour test)	Synaesthesia Battery Test	To understand if test scores differ between desktop and VR environments
VR Usability Study	Paper 2 / Chapter 5	Within-subjects A/B test	38	Wilcoxon Signed Rank Test, Repeated Measures ANOVA, Pairwise Comparisons	Open-ended feedback	Template Analysis	VR HMD (Forest vs Space virtual environments)	SUS, IPQ, CLM, SSQ	To explore immersive VR features that aid pitch-colour association engagement
Personalised vs Predefined Mapping	Paper 3 / Chapter 6	Between-subjects	30	Mann-Whitney U Test, ANOVA, Post-hoc Pairwise Comparisons	Open-ended feedback	Template Analysis	VR HMD (Space virtual environment)	Error Scores, Time Taken	To evaluate whether personalised or predefined mappings aid music theory learning

3.6 Ethical Considerations

All studies were approved by the University College Cork's Social Research Ethics Committee and conducted in accordance with ethical standards for research involving human participants (see Ethics Approval for Study One - Appendix A.3; Study Two - Appendix B.3; and Study Three - Appendix C.1). The ethics approval IDs can be found in Section 4.3.2 for Study One, Section 5.3.4 for Study Two, and Section 6.3.1 for Study Three. Participants provided informed consent (see Appendix A.1, B.1, and C.1) prior to taking part and were debriefed after each session. Measures were taken to ensure participant anonymity and data confidentiality throughout the research process.

3.7 Chapter Reflection

This methodology section provided a high-level overview of the methods used to help address the central research question presented in this thesis. The researcher adopts a pragmatic epistemological stance to guide the methodological approach to help understand the central question. Within this pragmatic standpoint, the constructionist viewpoint provides a more in-depth look at how the studies are designed with a balance of qualitative and quantitative approaches. The focus on creating immersive and interactive experiences that engage the participants when performing the pitch-colour cross-modal association task, aligns with the constructionist point of view of users being actively involved in the experience, learning by making.

The three studies presented in this thesis follow a clear path toward understanding how VR can enhance pitch-colour cross-modal associations. Study one (Chapter 4) explores VR's ability to be used to test pitch-colour associations tasks compared to a standard desktop set-up, seeking to answer the sub-question on VR as a viable technology for this type of perceptual task. The second study (Chapter 5) progresses the research to focus on a usability study exploring the different features of immersive VR settings that participants find beneficial for making pitch-colour associations. The final study (Chapter 6) positions the research in a use case scenario for learning basic musical exercises, exploring participants' performance when using a pre-defined pitch-colour mapping compared to a personalised relationship between the two senses.

This structure provides a transition from the validity of VR, exploring key immersive design elements, to testing a use case for the research area. The next chapter will present the pilot study seeking to explore the effectiveness of VR for understanding pitch-colour associations by comparing a consistency test with that taken on a desktop environment.

Chapter 4

Sound and Colour: Evaluating Auditory-Visual Tests in Virtual Reality and Traditional Desktop Settings

This chapter presents the study published as: P. O'Toole, M. Mancini, and I. Pitt. Sound and Colour: Evaluating Auditory- Visual Tests in Virtual Reality and Traditional Desktop Settings. *The International Conference on Advanced Visual Interfaces (AVI), 2024, pp. 1-5.* DOI: 10.1145/3656650.3656665. As this publication was a short paper, additional analysis, results and discussions are provided after the published content (see Section 4.7).

The following study explores the sub-question: *Is VR a viable platform for conducting pitch-colour association tests compared to desktop environments?*, that is presented in Section 1.3.

Abstract: *In this pilot study, we used a balanced A/B testing method to evaluate how an environment—a Virtual Reality (VR) head-mounted display (HMD) versus a traditional desktop—affects the performance of 20 participants on a musical pitch-colour association test. We aimed to discern the influence of testing environments on musical pitch-colour associations. Our findings revealed no significant difference in performance between the VR HMD and the traditional desktop conditions. Previous studies highlight VR's potential, but the results of our investigation—focusing on limited immersive and presence qualities—suggest that working in VR as opposed to a standard desktop environment has no significant influence on musical pitch-colour association test results. This outcome prompts a further investigation into other inherent VR characteristics, such as spatial audio, natural environment simulations, and interactive object manipulation, which could potentially enhance the effectiveness of auditory-visual association tests. This study is the first in a series of studies to explore how VR technologies can be exploited to augment multi-modal testing, with a fully immersive musical pitch-colour association test envisioned.*

4.1 Introduction

Virtual Reality (VR) has increasingly been embraced as an effective tool in research due to its ability to create immersive environments that simulate real-world experiences (Poguntke et al., 2019; Parsons and Phillips, 2016; Marto et al., 2022). One area where VR holds substantial promise is in exploring auditory-visual associations (Martin et al., 2022) and presenting synaesthesia-like experiences *(see Section 2.2.2)* to help understand connections between the senses (Curwen, 2018). Previously, pen and paper and desktop-based tests have been developed to determine if a person has synaesthesia (Asher et al., 2006; Carmichael et al., 2015). At the same time, these tests are also used to highlight the consistency of the general population in associating between their senses (Campen, 2010; Campen and Froger, 2003).

With VR and its immersive nature, a more intuitive and engaging context for these assessments could be created, allowing non-synaesthetes to experience multimodal stimuli in a way that more closely mimics the experiences of synaesthetes and possibly leads to enhanced consistency in pitch-colour associations within the test. However, the literature on the influence of VR on auditory-visual association tests is sparse and inconclusive.

This paper presents an experimental study designed to start filling this gap in the research. Utilising a balanced A/B testing methodology, we administered a musical pitch-colour association test to 20 participants using `The Synesthesia Battery`, an online platform designed for such assessments. Each participant completed the online test in two environments: a traditional desktop setting and a VR Head Mounted Display (HMD).

The research question (RQ1) was: *Does environment (Desktop vs VR HMD) significantly affect how participants perform in a musical pitch-colour association test?*

If the null hypothesis, stating that no significant difference exists between the test scores in both the desktop and VR HMD environments, cannot be rejected, we speculate that other elements unique to VR, such as spatial audio, natural environment simulations, and interactive object manipulation, might be the key factors in improving consistency in musical pitch-colour tests.

This pilot experiment will be the first step toward a broader understanding of how VR can be used to enhance multi-modal testing. By examining the impact of VR on pitch-colour association test performance, we aim to set a foundation for future research in this promising field.

4.2 Related Work

In this section a review of previous work on auditory-visual cross-modal associations, sound-colour synaesthesia, current auditory-visual association tests and related VR research is highlighted.

The aim is to contextualise our exploration of VR's potential to enhance understanding and assessment of auditory-visual associations.

4.2.1 Auditory-Visual Cross-Modal Associations

Auditory-visual association research has shown interesting connections between the auditory and visual senses. Lower-pitched sounds have been found to associate with darker colours, while higher-pitched sounds usually associate with lighter colours (Curwen, 2018; Ward et al., 2006; Goller et al., 2009). Other musical features also have associations with the visual sense, such as musical key-brightness, timbre-saturation, loudness-brightness and pitch-size (Palmer et al., 2013; Spence, 2011).

Emotions significantly influence how we perceive associations between the senses. Research on music-emotion and colour-emotion has revealed intriguing connections. Hevner's emotional model and Plutchik's wheel of emotions demonstrate the intricate relationships between emotion-music, and emotion-colour (Hevner, 1936; Plutchik, 2001). The emotion mediation hypothesis further suggests that individuals tend to select colours with emotional content similar to the music they listen to (Palmer et al., 2013). Understanding how emotions impact sensory associations is one of the factors to consider for designing innovative systems, especially in immersive environments like VR.

Research into synaesthesia has found that there are some similarities with cross-modal associations used in the general population (Ward et al., 2006; Williams et al., 2015; Lacey et al., 2016). Synaesthesia is a rare condition that affects approximately 2-4% of the population, with sound-colour synaesthesia affecting 0.2% of people (Curwen, 2018; Farina et al., 2017). It is a phenomenon that occurs when the stimulation of one sense modality gives rise to sensations in another sense modality (Curwen, 2018). Some people will associate the sound's pitch with a colour, shape, or size, but other aspects of music, such as timbre and musical key, can also associate with the visual sense (Curwen, 2018; Spector and Maurer, 2009; Brang and Ramachandran, 2011; Deroy and Spence, 2013).

With sound-colour synaesthesia, the experience is personalised and individualised for the person, and there is no single correct general auditory and visual match, and these matches are perceived either in the person's "mind's eye" or projected into the external space in front of them (Sagiv and Ward, 2006; Curwen, 2018).

4.2.2 Current Tests on Auditory-Visual Associations

Research on synaesthesia and auditory-visual associations has led to the development of numerous tests designed to determine whether an individual experiences a specific type of synaesthetic response or more generally how strong a person's associations are between the two senses. The current tests are `The Synaesthesia Battery` (Eagleman et al., 2007), `Test of Genuineness (TOG-R)` (Asher et al., 2006; Johnson et al., 2013), `NeCoSyn` method (Campen and Froger, 2003), and the `Implicit Association Test (IAT)` (Parise and Spence, 2012; Lacey et al., 2016).

`The Synesthesia Battery test` was selected as the test for this study, as it has a musical pitch-colour association test within its online battery of tests and has a researcher portal for easy management of participant scores (Eagleman et al., 2007). `The Synaesthesia Battery` test has been independently tested and validated, and is one of the most used tests in synaesthesia research to date (Carmichael et al., 2015). While it is used to identify certain synaesthetic abilities in participants, it also has the potential to test non-synaesthetes to understand the strength of their

associations between the auditory and visual senses (Campen, 2010; O'Toole et al., 2019).

4.2.3 Similar Virtual Reality Research

Although there is limited research on VR studies specifically focusing on auditory-visual association tests, related areas have utilised VR environments for similar tests called Stroop tests (Poguntke et al., 2019; Gradl et al., 2019; Wu et al., 2010).

The Stroop Test, a psychological assessment, illustrates the Stroop Effect, showing prolonged response times in colour naming compared to word reading. Some studies have found a higher reduction of stress in the VR version of the Stroop word-colour test compared to a regular desktop (Poguntke et al., 2019), with a follow-up study finding a 30-40% improvement in some metrics when using the VR version of the test (Gradl et al., 2019).

An overview on multi-modality in VR (Martin et al., 2022), highlights that 85% of multi-modality research in VR reports a positive impact, with only a 1% negative impact reported. Martin et al. (2022) states the potential for more research in cross-modality in VR, with limited research presented in the paper. In one study, it was found that cross-modal interactions between auditory and visual stimuli influence subjective experiences and assessments, including room evaluations in both real and virtual environments (Larsson et al., 2002).

Music, art, and synaesthesia have been sparse research topics when it comes to VR and cross-modal associations. However, studies on immersion and synaesthetic experience have identified several potentially immersive features, with audiovisual synchronisation and impressive graphics included (Silva Bastos et al., 2018). Other VR research shows musical timbre visual mappings (Soraghan, 2014), programs for learning piano Trujano et al. (2018), understanding of harmony in music Timoney et al. (2018), and creative expression Hosek (2023).

In the next section, an experimental study will be presented on comparing two environments when taking a musical pitch-colour association test.

4.3 Methodology

A pilot study was conducted to assess whether merely being present in a VR environment (rather than a conventional desktop setup) affects the way in which participants form pitch-colour associations. `The Synesthesia Battery`[1] online platform was used to conduct musical pitch-colour association tests in two VR scenarios (for details, see (Eagleman et al., 2007)).

The two environments being investigated used an online browser-based test, one in the more standard testing environment using a desktop computer and monitor, and the other being conducted on a VR Head Mounted Display (HMD).

Details on experiment design, participants, data collection, and analysis will be provided in the next sections.

[1] `https://synesthete.ircn.jp/home`



Figure 4.1: Testing environments used in Study One. Left: the desktop testing setup; Centre: the participant seating area for the VR test; Right: the VR testing environment using the Oculus Quest 2 headset.

4.3.1 Experimental Design

This study investigates musical pitch-colour associations using consistency tests on standard desktop and VR HMD setups. For the desktop, a PC with `Windows 10` and a `Phillips 246V5` monitor were used, set to full brightness and 6500k colour temperature. The VR setup employed an `Oculus Quest 2` HMD with `Oculus Touch` controllers, with the display manually calibrated to match the PC monitor's brightness. Both displays used Liquid Crystal Display (LCD) technology, with the monitor in `sRGB` and the HMD in `Rec.709` colour space [2] [3]. Participants used `Sony MDR-CD580` headphones, adjusting audio volume to their comfort. Spatial sound was not feasible due to limitations in the predefined `Synesthesia Battery` test environment.

The experimental design employed a within-subject A/B test. Half of the participants took the desktop test first, followed by the VR test, while the other half followed the reverse order. Participants heard 13 musical notes presented randomly 3 times each, totalling 39 auditory stimuli, covering a piano scale from middle C (C4 - 261.63 Hz) to one octave above (C5 - 523.25 Hz). They selected colours associated with the auditory stimuli using a colour picker (Figure 4.1). A five-minute break with an auditory-visual pattern video[4] helped mitigate potential memory effects from the first test (Asher et al., 2006). Participants completed surveys before the first test and after the second test. The experimental procedure, depicted in Figure 4.2, was followed by all participants, lasting approximately 30-40 minutes.

4.3.2 Participants

Twenty participants took part in the study, 80% were male, 15% were female, and 5% identified as non-binary, with 70% having little to no VR experience. The 25-34 age group was represented 11 times, while the 18-24, 35-44, and 45-54 age groups had 4, 3, and 2 representatives, respectively.

No participants reported having any issues with motion sickness before starting the study. Two participants stated that they had synaesthesia, but it was not musical pitch-colour synaesthesia that is being tested in this study.

[2] `https://developers.meta.com/horizon/resources/color-management-guide/`, accessed on February 8, 2025

[3] `https://developers.meta.com/horizon/resources/color-brightness-mastering/`, accessed on February 8, 2025

[4] `https://youtu.be/k4FCoB30RNA`

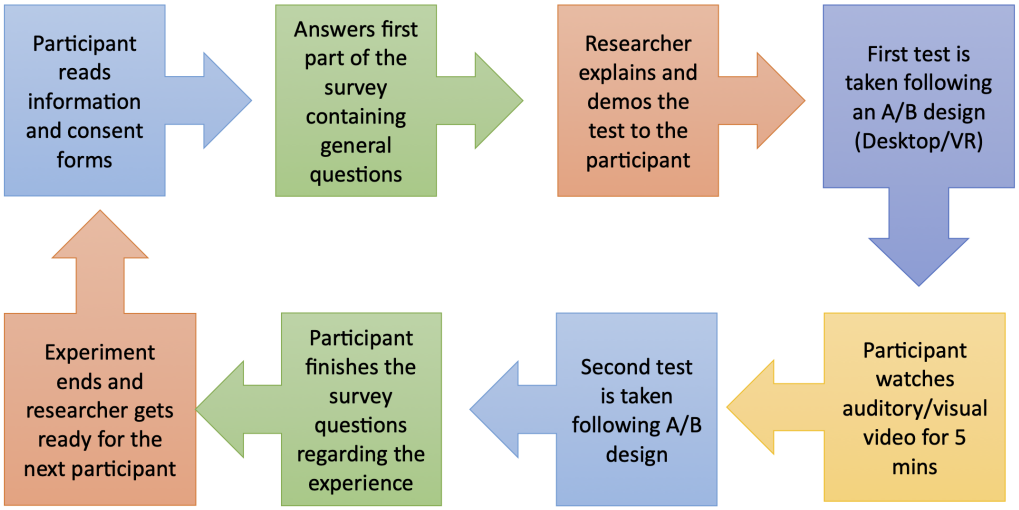


Figure 4.2: Study design for Study One: participants followed the steps in a clockwise sequence starting from the top left. The entire testing session lasted 30-40 minutes per participant.

Ethics approval was granted by the UCC Ethics Review Board (Log 2023-025), and the study was conducted in accordance with the Declaration of Helsinki.

4.3.3 Data Collection

`Google Forms` collected general demographics and insights on VR experience and testing environments. Data from musical pitch-colour tests were obtained via `The Synesthesia Battery` website, with a dedicated researcher's portal allowing data export as `Comma-Separated Values (CSV)` files. To limit personal data collection, a single account was used to collect all participants' tests, with participants providing their test IDs in the survey form to keep track of which environment was associated with which test. `CSV` files contain the average score used to measure sound-colour associations (with synaesthetes deemed < 1.0 and $1.0 >$ deemed as non-synaesthetes).

4.3.4 Data Analysis

`SPSS` was used to analyse the question stated in **RQ1**; with a Paired T-Test being used. From the insights participants shared during the study, a few other questions were analysed to gain insights that will help develop future studies. A Fisher's exact test for both VR experience and sound-colour association difficulty was performed to gain further insights, as a chi-squared test might not be useful with sample sizes under 50 (Greene and D'Oliveira, 2011). Analysing participants' feedback will help design future tests and VR applications providing associations between auditory and visual senses.

4.4 Results

To answer **RQ1**, stated in Section 4.1, a Paired T-Test was used. Results from the Paired T-Test show a mean difference of *-0.06051650* between the desktop and VR test scores, with a significance level of *0.605* (non-significance) on the two-sided p-value. This result does not allow us to reject the null hypothesis meaning that no significant difference could be found in the test scores between the desktop test environment and the VR test environment.

For additional insights, a Fisher's exact test was conducted to see, firstly, if participants' VR experience and, secondly, their difficulty associating sound to colours had any significant effect on participants' preferred testing environment. For the VR experience, Fisher's exact test shows a significance of *0.010*, which indicates that VR experience plays a role in the participants' choice of the preferred testing environment. For the difficulty associating sound and colour stimuli, Fisher's exact test shows a significance of *0.358*, which reveals no significant association between the preferred testing environment and the difficulty of the task performed in the test.

4.5 Discussion

The answer to **RQ1** leans towards the null hypothesis, confirmed by the Paired T-Test results, suggesting no significant difference in test scores between the two environments. This study represents an initial exploratory step towards understanding the impact of VR on auditory-visual associations.

While 12 out of 20 participants felt more comfortable associating the musical notes with colours in the VR environment, the results show no significant differences between the two testing environments in this A/B study design. Given that simple presence in a VR environment—with minimal interaction or immersive features such as spatial audio other than moving one's head and use of controllers—does not affect musical pitch-colour associations, we now have the opportunity to investigate other immersive features unique to VR. These further explorations may help enhance our understanding of the interplay between the two senses.

To explore the participants' perception of their performance in the two environments and see if there was any connection between how difficult they found associating sound, the colour stimuli, and the participant's VR experience, a Fisher's exact test was used. As reported in Section 4.4, the test shows that there was no significant correlation between participants' choice of the preferred test environment and how difficult they found it to associate between their auditory and visual senses.

Although participants expressed a preference for the `Virtual Reality` environment, **RQ1** revealed no significant difference in performance between the testing environments. Interestingly, participants' perceptions of their performance were not aligned with the difficulty they experienced in associating sound and colour stimuli. This shows that self-reporting tests can be flawed, and while you can gain interesting insights, backing it up with rigorous tested scientific process is important (Eagleman et al., 2007; Asher et al., 2006).

Analysis using Fisher's exact test highlights a significant correlation between participants' prior VR experience and their preferred testing environment. Interestingly, 70% of participants had minimal to no VR experience before this study,

potentially influencing their environment choice. This suggests that the novelty of VR immersion may sway participants' preferences. However, as VR becomes more ubiquitous, the initial novelty effect wanes, although it remains a promising tool for training and testing (Wiesing et al., 2022). In the context of auditory-visual association tests, the standard desktop test lacks realism in sound perception. Placing participants in immersive VR environments with high-quality spatial audio could foster more natural sensory associations.

4.5.1 Implications & Limitations

The study reported here found no significant difference between using a standard desktop and a VR HMD when participants were tested on how they form associations between their auditory and visual senses. It provides no support for the hypothesis that using a basic VR environment rather than a desktop environment affects auditory-visual associations and paves the way to investigate other features of VR and how they can be used to improve and strengthen cross-modal associations between our two senses. Further investigation into features such as spatial audio, realistic environments, interactions with objects and the ability to manipulate the size and shape of objects could provide fruitful insights. Research has shown that people associate not just colour brightness, but other visual aspects like saturation and size with auditory features, such as timbre, loudness, and pitch (Spence, 2011). These aspects will be the focus of our future studies in auditory-visual associations using VR.

Participants taking both testing environments in the same session is not ideal and while some previous studies have done this, it can lead to memory effects between the two tests. To mitigate potential memory effects mentioned in Section 4.3.1, participants watched a video during the break, however having separate sessions weeks or months apart would solve any potential issues. Conducting all tests in a single session was chosen to increase participation numbers.

Assessing sound-colour associations in different environments, particularly VR, presents challenges due to device-specific colour management limitations [5] [6]. Calibrating colours on both devices was impractical for the VR headset, necessitating steps to ensure consistent colour presentation, including using a colour chart on both displays. Both devices utilized LCD screens with standard colour temperature settings. Prior research underscores the imperative for standardised colour management in VR environments (Billger et al., 2004; Díaz-Barrancas et al., 2020; Gil Rodríguez et al., 2022).

4.6 Conclusion

This pilot study delved into the potential impact of testing environments, namely a VR head-mounted display (HMD) and a traditional desktop setup, on the performance outcomes of a musical pitch-colour association test. A balanced A/B testing approach involving 20 participants was employed for the investigation. Our primary

[5] `https://developers.meta.com/horizon/resources/color-management-guide/`, accessed on February 8, 2025

[6] `https://developers.meta.com/horizon/resources/color-brightness-mastering/`, accessed on February 8, 2025

research question, *"Does environment (Desktop vs VR HMD) significantly affect how participants perform in a musical pitch-colour association test?"* found its answer leaning towards the null hypothesis. The outcomes revealed no significant difference in the performance of participants under the VR and desktop conditions.

Given these findings, our study could not confirm the alternative hypothesis, demonstrating no significant difference in test scores between the two testing environments. These outcomes direct our future investigations toward a more immersive VR application. Recognizing VR's transformative potential in similar research domains (see Section 4.2.3), it becomes crucial to understand the factors contributing to the effectiveness of auditory-visual association tests in VR settings.

4.7 Delving Deeper into Auditory-Visual Tests in VR & Desktop Environments

The previous sections of this chapter are reproduced as published in the short paper accepted to the *The International Conference on Advanced Visual Interfaces (AVI).* The remainder of this chapter delves deeper into the comparison between the two environments when conducting auditory-visual association tests. Further quantitive analysis on colour brightness and distance is explored, while a qualitative examination using a template analysis is also presented. This deeper investigation provides more insights on the use of VR HMDs for exploring sound-colour cross-modal associations.

4.8 Additional Data Analysis

In addition to the primary analysis presented in Section 4.3.4, further explorations were undertaken to gain additional insights in relation to the pitch-colour task and participants experience of the tests that could inform future study design. The two additional analyses are presented below.

4.8.1 Colour brightness & distance choices

First, a brightness value was computed for each selected colour using the `International Telecommunication Union Radiocommunication (ITU-R) BT.709`[7] perceptual luminance model, which weights RGB channels according to human visual sensitivity. This allowed for an investigation into whether participants tended to associate higher-pitched notes with brighter colours. Spearman's correlation analyses (non-parametric data) were conducted separately for each condition (VR and Desktop) to examine the relationship between pitch height and perceived brightness. These analyses provided additional insight into cross-modal congruency patterns across both environments.

Second, the consistency of colour choices for each musical note was assessed by calculating the average Euclidean distance between the RGB values selected by participants. This provided a measure of agreement across the group: lower average

[7] `https://www.itu.int/dms_pubrec/itu-r/rec/bt/r-rec-bt.709-6-201506-i!!pdf-e.pdf`

distances indicated greater consistency in colour association for a given pitch. This analysis was carried out independently of test environment and was intended to reveal general patterns in participant mappings.

4.8.2 Qualitative Analysis

Finally, qualitative data from the post-study questionnaire were analysed. `Voyant Tools`[8], a web-based text analysis platform, was first used to explore participants' responses. This initial text analysis helped identify common words and how they were linked, which in turn highlighted recurring themes. These themes informed the development of the *"a priori"* codes used in the subsequent template analysis. Together, the text analysis and template analysis provide insights that can guide future design decisions for research on cross-modal pitch-colour associations and their implementation in applications, especially within immersive VR contexts.

4.9 Additional Results

4.9.1 Colour brightness & distance choices

To explore whether participants associated higher musical pitches with brighter colours, a brightness metric was calculated from the `RGB` values selected in the pitch-colour association test. This used the `ITU-R BT.709` perceptual luminance model, which weights `RGB` channels according to human visual sensitivity:

$$Brightness = 0.2126 \times R + 0.7152 \times G + 0.0722 \times B \tag{4.1}$$

Separate analyses were conducted for each condition (VR and Desktop), based on participant IDs mapped to their respective environments. In the VR condition, results showed a moderate positive correlation between pitch height and brightness values (Spearman's $\rho = 0.511$, $p < .0001$), suggesting that higher-pitched notes were generally associated with brighter colours. In the Desktop condition, the correlation was slightly stronger (Spearman's $\rho = 0.539$, $p < .0001$), again supporting the trend of higher pitches being matched with brighter colours.

Visual inspection of scatter plots (see Figure 4.3) confirmed this relationship in both environments, with average brightness increasing consistently as pitch rose.

Also, the colour distances for each note in the test (ranging from C4 [261.63 Hz] to C5 [523.25 Hz], based on the `A440 or ISO16` tuning standard) were explored to determine whether or not participants showed any similarities in their musical pitch to colour choices. The calculation was performed to find the notes with the least distance in their `RGB` values. While a `Hue, Saturation, and Value (HSV)` colour picker tool was used for the test and is more intuitive to the way people think, the files outputted from the online test save the colour data as `RGB` and are computed as such in the analysis. For each note, we considered the set of colours chosen by all participants. Then, between each pair of colours $(C1, C2)$, we computed the following distance:

$$d(C1, C2) = \sqrt{(C1_R - C2_R)^2 + (C1_G - C2_G)^2 + (C1_B - C2_B)^2} \tag{4.2}$$

[8] `https://voyant-tools.org`

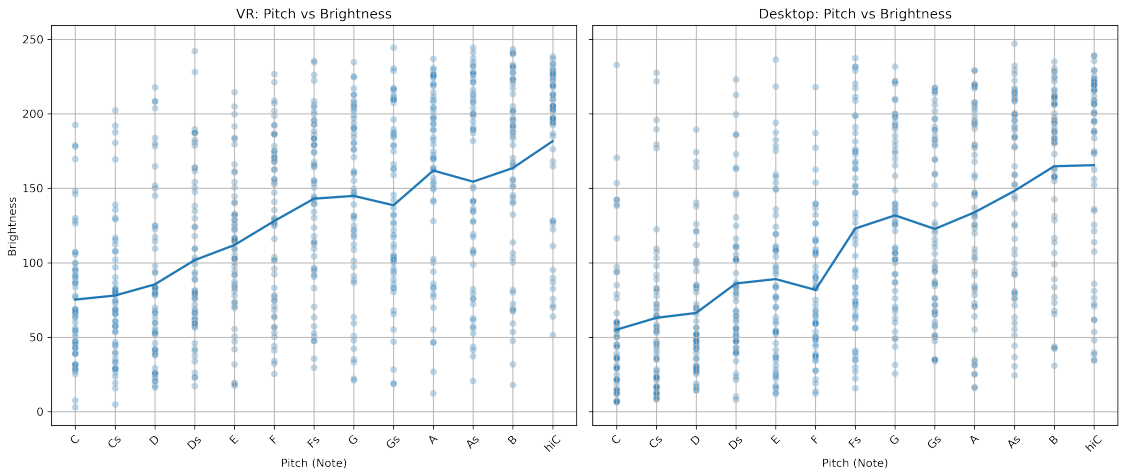


Figure 4.3: Left: Brightness levels chosen for each note in the VR setting, with the trend line showing the average brightness; Right: Brightness levels for each note in the Desktop setting, with the trend line highlighting the average brightness for each note.

where (C_R, C_G, C_B) are the `RGB` components of colour C. Finally, we averaged the distances between all the colour pairs for that note. Results showed that the note C4 had the smallest average distance between participants' colour selections (see Figure 4.4), indicating a greater consistency in how this pitch was perceived in terms of colour.

4.9.2 Post-Study Questionnaire Analysis

Participants were asked which test environment they felt was more conducive to making pitch-colour associations in, with 60% choosing the VR test and 25% choosing both (see the image on the right of Figure 4.5). In Figure 4.5 (the pie chart to the left), it also shows that 65% felt that the both testing environments gave a similar experience when conducting the pitch-colour task.

In response to the question, *"How difficult did you find the task of associating colours to musical auditory stimuli?"*, participant responses were fairly evenly distributed across the middle difficulty levels, with the most frequent selections being *"easy"* (6), *"ok"* (5), and *"hard"* (7) (see right-hand image in Figure 4.6).

Using `Voyant Tools` to analyse the key terms mentioned by participants from the open-ended questions in the post-study questionnaire revealed some interesting connections. Figure 4.7 highlights strong links between the term *"VR"* and the terms, *"focus"*, and to a lesser extent, *"immersive"*, *"experience"*, and *"easier"*. Also some of the other terms mentioned most (see Figure 4.8, were *"controllers"*, *"intuitive"*, *"mouse"*, and *"environment"*.

From this text analysis, initial *"a priori"* codes where developed for the template analysis. These initial codes centred and *"focus"*, *"immersion"*, *"experience"*, *"difficulty"*, *"controllers"*, and *"environment"*. After further exploration of the data using `NVivo`, three main themes were decided on, with *"Environment"*, *"Experience"*, and *"Interaction"* chosen (see figure in Appendix A.6). These themes were used to explore the participants' feedback on the two open-ended questions asked of them in the post-study questionnaire. The results of the template analysis is discussed below with a section for each of the three main themes highlighted.

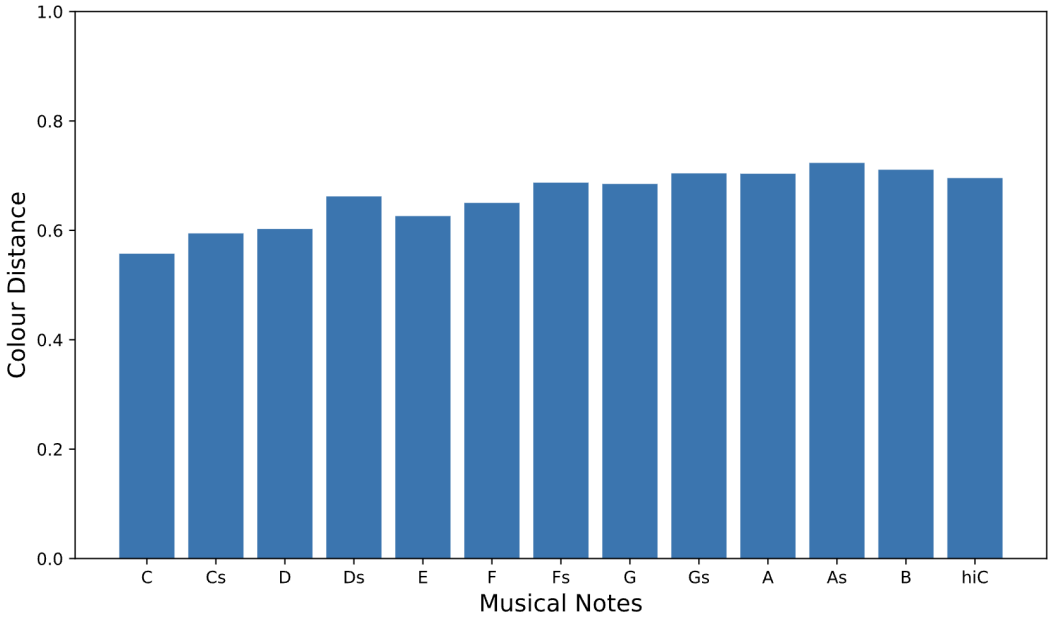


Figure 4.4: Overall colour distances for each musical note in the scale from C4 to B4, including a high C note (*"s"* = Shape and *"hiC"* = high C or C5 note). For each note, the set of all colours chosen by participants was collected, and the average distance between all colour pairs in the set was calculated. The results are normalised between $[0, 1]$, where 0 indicates that all participants selected the same colour for a given note, and 1 indicates that they chose maximally different colours (e.g., black vs. white).

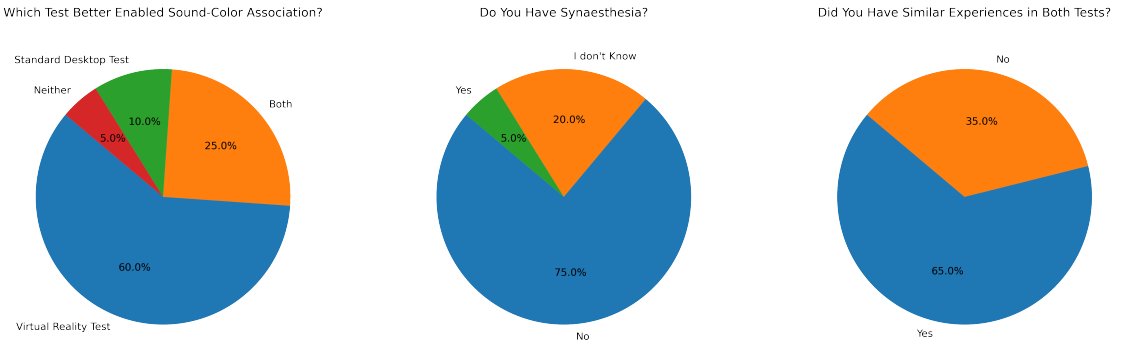


Figure 4.5: Left: Participants' choices of which test environment they felt best enabled them to make pitch-colour associations; Centre: Distribution of participants who reported having synaesthesia; Right: Participants' responses on whether they felt the test experience was similar in both environments.

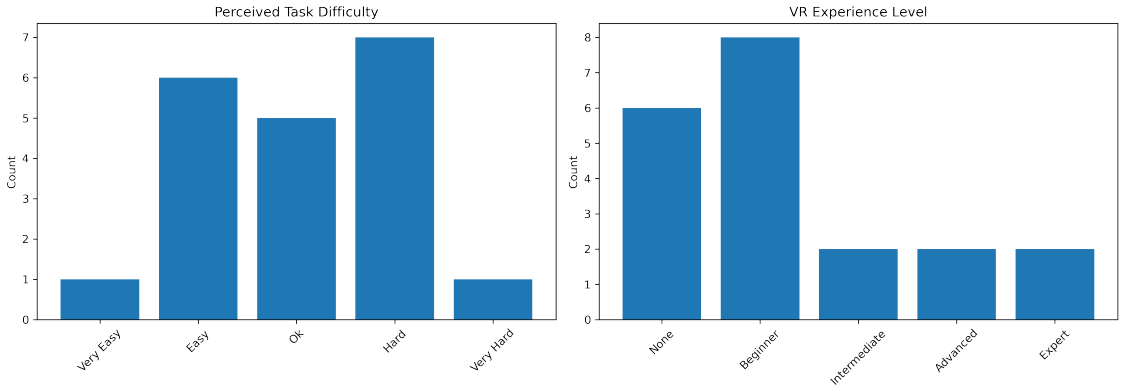


Figure 4.6: Left: Perceived difficulty of the pitch-colour association task, irrespective of environment; Right: Distribution of participants' levels of VR experience.

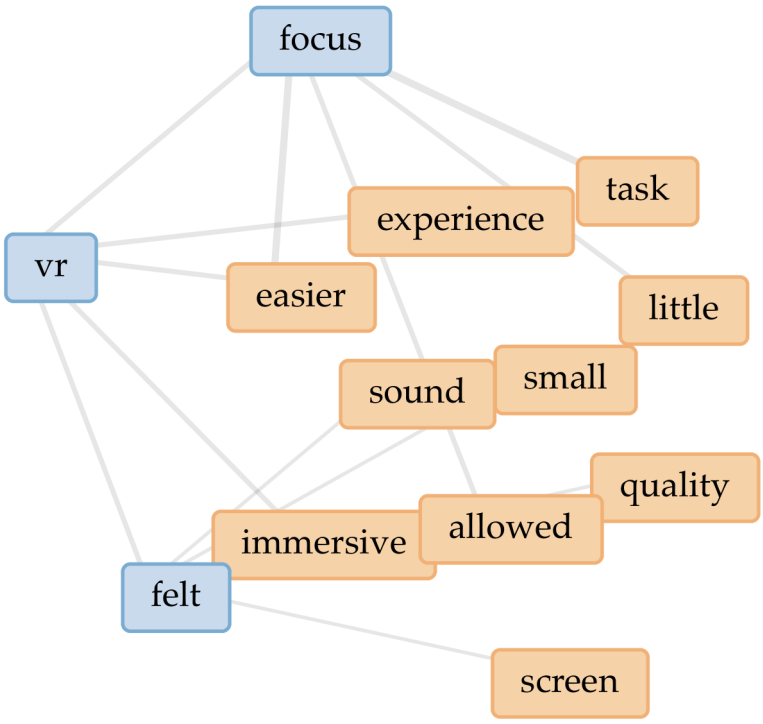


Figure 4.7: Text links showing the words most commonly mentioned by participants when describing preferred aspects of the Virtual Reality experience compared with the desktop environment. Blue highlighted words indicate a strong link between two words, while orange highlighted words show a slightly weaker link.



Figure 4.8: Word cloud illustrating participant feedback on preferred aspects of the VR experience compared to the real-life environment.

Theme One: Environment

The first theme investigated centres around environment, with the two subthemes of *"focus"* and *"audio perception"*. When asked the question: *"What, if any, aspect of the Virtual Reality experience did you like or prefer to the test in the real life environment?"*, some interesting insights were found regarding participants level of focus and how they perceived the auditory stimuli in the VR setting compared to the desktop.

Focus: With regard to *"focus"*, participants found it was easier to focus and interact with the task in the VR setting. P20 stated, *"Less outside distraction, easier to focus on the task"*, while P15 mentioned that there was *"less visual distraction in virtual reality i.e., easier to focus on colours"*. They also mentioned (P15), *"I felt like I didn't need to focus on picking a colour in VR"*, when describing why they preferred the VR environment for the task. This highlights several potential benefits of using VR for perceptual tasks such as pitch-colour association. The ability to minimise external distractions and fully control the visual environment can enhance focus and reduce cognitive effort. By immersing participants in an engaging, distraction-free setting, VR allows users to concentrate more naturally on the task itself, without the added burden of consciously trying to maintain focus.

P19 also found the VR setting provided more focus even though the test interface was the same in both conditions, stating *"you feel more isolated, and focused. But because the interface was similar I did not had completely different feeling. Maybe if it was a bit different than a webpage in the vr it was going to be different"*. This highlights that the closed-off environment of the VR HMD can isolate and therefore focus the participants, but with the same browser-based test in both settings the feelings of the environment can be minimal, without designing for an immersive VR space.

Audio Perception: From the post-study feedback, audio was mentioned by several participants. When talking about the test in the VR environment, P13 mentioned that *"It felt a small bit more immersive when listening to the audio in VR"*.

A similar sentiment was stated by P19, when they said *"The sound quality seemed better. I felt like I was in a much bigger room"*. These are interesting insights because the audio design was similar for both environments, as it was a web-based pitch-colour association test. There might be slight differences in how the audio is processed between the two different devices, but from the statements above the audio differences seem to be more perceptual. P5, who answered *"No"* to having synaesthesia, stated that *"The fact that the sound is present visually more in a VR system, whereas looking at a PC Screen"*. This is another interesting insight, as the visuals were similar. The only notable difference between the two test environments was the size and shape of the browser window, which was more square-shaped. However, the main testing area on the desktop monitor was also square, with additional blue side-panels used to fill the wider screen space (see Figure 4.1). These insights on the perception of the audio stimuli in the VR environment and how the sound was perceived as more visual highlight some interesting points to contemplate.

In summary, the VR environment provided some interesting insights. A high number (n=7) of participants mentioned how they could focus more on the pitch-colour interactions using the HMD. In addition, the perception of the auditory stimuli was different between the two environments, even though the same browser-based test was used in both. While the VR HMD processes audio slightly differently depending on the user's head movements, this difference is mainly noticeable when the user is not directly facing the browser screen in front of them, and should not affect the perception of the sound as participants' described. This may suggest that increased immersion and focused task engagement can perceptually alter how the audio is experienced. In the next theme, *"experience"* is discussed, looking at the role of immersion and the pitch-colour task had on participants.

Theme Two: Experience

As highlighted in the Section 4.9.2, strong clusters around the word *"VR"*, were found with the terms *"immersive"*, *"easier"*, and *"experience"*. These terms (immersive and experience) would be synonymous with VR, and with 70% of participants stating that they had minimal or no exposure to VR, it could be that it was a new experience and a level of immersion that they have not felt before. With this theme, the subthemes of *immersion* and *pitch-colour test"* are explored to understand how the VR experience (minimal immersive features) impacted participants in taking the task.

Immersion: The VR HMD, which was the preferred environment for the majority of the participants, was unexpectedly deemed as immersive even though it was presented in a dark simple space with a web browser forward-facing. P3, who never used VR before, stated: *"I found my experience in VR more immersive and easier to focus on the task. Making associations during my VR experience seem more natural and fluid to me than the real life environment"*. Other participants stated that they found the VR test *"more interactive"* (P1) and *"more intuitive"* (P12). These statements highlight some potential reasons for the increased immersion in the VR HMD setup. As mentioned previously in the section on audio perception, differences were acknowledged between the two environments. P13 mentioned: *"It felt a small bit more immersive when listening to the audio in VR"*, while P9 stated,

"Using the controllers made the experience feel more immersive". The two potential differences in the set-ups were the controller interactions (which will be discussed in the last theme) and the perception of the presented audio. While these were not designed in anyway to increase the immersion, the standard use of these element in the HMD managed to provide an immersive experience to multiple participants.

Pitch-Colour Test: The order of the pitch-colour test also seemed to impact how the participants felt about taking the task. P6 mentioned that it was *"hard to articulate, just seemed more drawn in and decisive in VR. If VR was done 1st it may have been the opposite"* when discussing why they preferred conducting the test in the VR environment. On the other hand, P18 stated, *"I feel it was more easier/better on the desktop not because the environment but because I have already did it one before using the VR"*. The second environment, irrespective of which one it was, might play a role in the participants' preference and perceive difficulty of the task.

P16 mentioned, *"There were less stimuli in the [VR] environment - most of my vision was occupied by the colour grid so it helped focus and streamline my thoughts about colour once I heard the music note. Outside of VR, I felt slightly more distant from the test"*. This further highlights the impact VR had on participants' ability to focus. Additionally, the insights on the participant feeling *"more distant"* when taking the desktop test suggests that VR provided not only fewer distractions but a perceived closeness to the task in the VE. The enclosed space of the VR HMD might make the distance seem less, especially with no other objects to focus on in the environment. Providing adaptable parameters for the task, such as distance from the colour palette might improve the usability and comfort for users when conducing the test.

The subthemes from the *"Experience"* theme revealed unexpected increases in immersion within the VR HMD in some participants (n=6) irrespective of which environment was taken first, despite its low-immersion design. Differences in interaction style and audio perception were highlighted as potential contributing factors. Additionally, the positioning of the VR test in the A/B study design impacted participants' experience of the environment, which the second test improving the decisiveness and difficulty of the task. The final theme will explore *"Interaction"*, focusing on the subthemes, *"controllers"* and *"engagement"*.

Theme Three: Interactions

The third theme that was generated while conducting the template analysis revolves around interactions. Participants mentioned how VR's interactivity and intuitiveness help with the pitch-colour task. This theme was divided into two subthemes: *"engagement"* and *"controllers"*. The following section discusses each subtheme in detail, supported by participant feedback.

Engagement: Participants found that taking the pitch-colour test in the VR environment engaged them more in the task. P16 stated, *"I felt more involved with the colour grid - found it a little easy to associate colours when in VR"*. The more interactive nature of VR seemed to immerse the participant helping with the understanding of the associations between the auditory and visual senses. To support

this statement, P9 also found the interactive and embodied nature of the VR setting helped with immersion, presence and focus. They (P9) stated that *"The immersive nature of the test and having to concentrate on what I was doing with my hands made me feel more present during the test"*.

Even in a low-immersion VR environment with no additional immersive features, participants stated having increased engagement with the task, which added to the experience as well as helping with focus on the pitch-colour task. Controllers were the main interaction feature that helped improve the participant's experience. The next subtheme will explore the benefit of controller interactions when conducting the pitch-colour task.

Controllers: In terms of interaction in the VR setting compared to the desktop, several participants mentioned the VR controllers as one of the aspects that helped them with the task. When asked what they preferred in the VR task compared to the desktop, P11 stated that it *"took slightly less time to point and select the colour in the VR test as the controllers were more intuitive than the mouse"*. Similarly, P10 mentioned, *"I find pointing through VR much easier than scrubbing with a mouse meaning I feel I can get a more accurate colour selection"*, and P9 highlighted that *"using the controllers made the experience feel more immersive"*, from an interaction standpoint the VR setup seemed to provide multiple benefits to participants. Participants described the VR setting as offering increased immersion, quicker and more intuitive movement, greater engagement, and improved accuracy during the test.

To summarise the third theme, levels of engagement were increased in the VR environment which participant found helped them engagement and focus on the task. The controller interactions were mentioned frequently by participants as a feature that helped immerse them in the task and that also were more intuitive and efficient than the interactions in the desktop setting. From these results, the next section will look to find meaning in how participants experienced the two testing environments and seek to answer **RQ1**, while also exploring the additional results to help inform the future direction of this research.

4.10 Additional Discussion

Section 4.4 discusses the findings in relation to **RQ1**. This section provides additional insights on the research question presented in Section 4.1, as well as insights into pitch-colour choices made by participants and more broader insights on potential directions for improvement, and supplementary observations drawn from the data.

4.10.1 Effect of Environment (Desktop vs VR HMD)

While participants scored similarly across both conditions, suggesting that merely conducting the existing online test in a low-immersion VR environment does not inherently impact performance. Results for the subjective feedback indicated some participant preference towards VR (12 out of 20 participants felt more comfortable in the VR environment), suggesting a potential avenue for further exploration into user experience rather than pure performance. This aligns with findings by McMahan

et al. (2006) that suggest the effectiveness of VR is highly dependent on a multitude of factors, such as interaction design, rather than immersion alone. Although the VR setup was intentionally designed with limited immersion, certain elements—such as the control interactions and the enhanced focus provided by the VR environment—appeared to benefit participants. These factors will be further examined in Section 4.10.2, which draws on participant feedback to identify ways the study could be improved in future iterations.

With the findings regarding pitch-brightness presented in Section 4.9.1, this confirmed previous research on cross-modal associations where lower pitch and darker brightness and high pitch and higher brightness were found (Ward et al., 2006; Spence, 2011). This suggest that while the immersive environment did not significantly impact overall test scores, participants in both conditions demonstrated a similar perceptual mapping between pitch and colour brightness. While the previous research on pitch-colour cross-modal association was conducted in real world setting, this study is potentially the first to confirm these connections in VR environments. Malpica et al. (2020) found the presence of auditory-visual cross-modal associations more generally in a VR setting, however their exploration did not include the modality features of pitch and colour. Now that it is known that pitch-colour cross-modal associations are present in VR settings, can the understanding between the two modality features be strengthen by harnessing the full potential of the technology?

Another observation highlighted in Section 4.5, regarding the impact of testing environments concerns the relationship between participants' prior VR experience and the choice of environment they perceived to be more effective in when making pitch-colour associations. Results from the Fisher's exact test shows a significant association between the chosen environment and the extent of previous VR experience. When asked about their prior experience with VR, six participants reported having no experience, while eight described their experience level as beginner (see the right image in Figure 4.6). This preference among inexperienced VR users for the test may be attributed to the excitement or *"wow"* factor of encountering a new experience, such as using a VR headset for the first time. Being present even in a simple VR environment can play a significant role in how participants chose their preferred testing environment. Previous VR studies have have found that the novelty effect can impact results of studies (Sinlapanuntakul et al., 2022; Wiesing et al., 2022). To reduce the impact of the novelty effect, Ramírez-Sanz et al. (2023) introduced a VR tutorial space before participants began the main study. This approach could be a great way to teach participants what is needed to conduct the test in VR while also getting them used to the environment before taking the test.

Next, we examine in greater depth the findings reported in Section 4.5 concerning the association between perceived task difficulty and participants' preferred testing environment. As reported in Section 4.4, the test shows that there was no significant correlation between participants' choice of the preferred test environment and how difficult they found it to associate between their auditory and visual senses. With the majority of participants choosing the VR setting as the preferred environment (60%), and the difficulty evenly spread between the three middle options of *"easy"* (6), *"ok"* (5), and *"hard"* (7), environment did not help participant in how the perceived the difficulty of the task. This highlights the need for further exploration into using a more immersive VR environment to make pitch-colour associations.

As previously highlighted, added immersion, interaction, and presence have

shown improved performance in multi-modal task in VR settings (McMahan et al., 2006; Dede, 2009; Marucci et al., 2021). Using VR's capabilities to engage and immerse the user in a more meaningful way, using interaction techniques, 3D audio and visual elements could possibly help play a part in creating a better understanding of how a person associates cross-modally using pitch and colour.

From seeking to answer *RQ1*, it was found that there was no significant difference in test scores between the two environments, leading to believe a low-immersion VR setting can perform at a similar level to the standard desktop test. Also, pitch-colour cross-modal associations were found in the VR setting, showing a similar relationship to the connections made in the desktop environment. Elements such as novelty of VR, and task difficulty, irrespective of environment was discussed with using more immersive features of the technology an interesting next step. Future improvements are discussed next in more detail, taking from the template analysis on the themes of *"Environment"*, *"Experience"*, and *"Interaction"*.

4.10.2 Insights for Future Improvements

The qualitative analysis of the post-study questionnaire provided valuable insights into how VR could be used more effectively to deliver pitch-colour association tests and strengthen the cross-modal relationship between auditory and visual modalities. As discussed in Section 4.9.2, the three themes of *"Environment"*, *"Experience"*, and *"Interaction"* revealed specific VR features that may enhance participants' perceptual understanding of pitch-colour associations.

From the *"Environment"* theme, *"focus"* and *"audio perception"* were mentioned frequently by participants. It was interesting that *"focus"* was mentioned so frequently as a factor in the preference for the VR setting. The VR environment was not very different to the desktop setting, with a dark room with a browser facing forward; however, the HMD seemed to focus participants and provide a more engaging test.

If a low-immersion VR setting can provide this level of focus compared to a desktop environment, it could be beneficial in many learning and training areas. Previous studies have shown how different VR environments can affect mood, focus, and memory (Gkoumas and Shimi, 2022; Yu et al., 2018). Also, if focus is an important factor to provide for the intended task, studies have shown that immersion and presence can be increased without the use of visually rich environments, but through thoughtful multimodal design (Potter et al., 2022).

"Audio perception" emerged as an interesting theme in the template analysis, highlighting the way sound is designed can influence how participants feel in a virtual environment. While there was no control over how the audio was played during the online test, the `Quest 2` headset presents basic spatial audio effects even on the home screen, where the sound shifts slightly due to head movements.

Some participants noticed the change in spatial audio and mentioned that the way the audio was presented helped them feel more immersed. The feedback around audio also suggested that this wasn't just about immersion, but perception. P5 (see Section 4.9.2) noted that the sound felt more *"visual"* in the VR environment, which could be understood as a more heightened perception of how the auditory stimuli associates with the visuals presented. This could highlight VR's ability to integrate multiple senses and support stronger connections in sound-colour cross-

modal associations, which is the aim of this research.

If participants are already starting to form meaningful associations between pitch and colour in a low-immersion VR test with basic 3D audio, this raises the question of whether adding more immersive features could enhance those associations. Features such as fully 3D spatial audio within a fully interactive environment may make those associations even stronger. As highlighted in previous research, spatial sound in VR can increase immersion, presence, and even performance (Potter et al., 2022; Broderick, 2021; Gröhn et al., 2005; Huang et al., 2024).

Placing participants in more realistic virtual spaces with high-quality spatial audio may offer a more natural way for them to form links between sound and vision, something that current desktop-based tests do not really provide.

The frequent mention of *"immersion"* in participants' feedback about the VR test suggests that immersiveness is not solely dependent on realistic visuals, but also on other factors within the virtual environment. While some definitions of immersion emphasise visual fidelity—often referred to as technical immersion—other feature modalities, such as sound and interaction, are equally important in contributing to a fully immersive experience (Bowman and McMahan, 2007; McMahan et al., 2006; Grimshaw-Aagaard and Walther-Hansen, 2024). This study has highlighted that, without much visual complexities in the VR environment, participants experience immersiveness through other features, such as controller interactions and auditory presentation. This shows that high quality realistic graphics are not needed to immerse users, and with multimodal design using the other modality features, such as 3D audio, similar immersion can be obtained (Potter et al., 2022; Chalmers et al., 2009).

The interactions in the VR test, especially with the controllers, were highlighted as a feature that engaged, immersed, and felt more natural for the participants (see Section 4.9.2). Studies have found that interactions in VR can affect immersion and presence for the user (Regenbrecht and Schubert, 2002; Padhi and Katoch, 2021), while it has also be found the more realistic or intuitive interactions can benefit performance when conducting tasks in VR settings (McMahan et al., 2006).

The VR test used a ray-based interaction system, allowing participants to point and select elements on the screen using the controllers. This contrasts with the mouse-based interaction in the desktop version. The use of two motion-tracked controllers in VR enabled more natural and intuitive engagement, offering a significant advantage over the keyboard-and-mouse setup of the desktop environment. McMahan et al. (2006) found that, when comparing three different interaction techniques, the method based on 2D mouse input resulted in significantly longer object manipulation times than the other two 3D-based techniques.

Well-designed interaction in VR can do more than engage users; it can enhance immersion, support performance, and contribute to a stronger sense of presence. Looking ahead, a key challenge for future VR-based pitch-colour research is to design interactions that feel intuitive while helping users form and reinforce cross-modal associations in a more natural way. Building on the earlier discussion of potential improvements for presenting pitch-colour tasks in VR, the following section explores additional insights into musical pitch perception, colour associations, and synaesthesia, including participants whose scores closely aligned with those typical of synaesthetes.

4.10.3 Additional Observations:

With the low C note showing the least amount of distance in colour choices across all participants (see Figure 4.4), this suggests a higher level of agreement in how it is perceived and mapped to colour compared to other notes. One possible explanation is the widespread familiarity with the note C4, commonly known as *"Middle C"*. This note is often the first reference point learned by beginners on the piano, and may hold a subconscious salience due to early musical exposure. Supporting this interpretation, Janata et al. (2007) found that familiar music is capable of triggering vivid recollections of personal experiences, particularly from childhood. Given that the stimuli in this study consisted of piano notes, it is plausible that the C4 note evoked memories or learned associations in participants who had early exposure to piano learning.

Additionally, C4 is the lowest-pitched note included in the test, which may have influenced how it was perceived in colour space. Research has shown that lower pitches tend to be associated with darker colours, while higher pitches are linked with lighter or brighter hues (Curwen, 2018; Ward et al., 2006; Goller et al., 2009). From this perspective, the lower and higher extremes of the tested range may provide stronger perceptual anchors, making them easier to associate with consistent colours. In contrast, mid-range pitches may lack this perceptual anchoring, resulting in greater variability in colour association across participants (O'Toole et al., 2019).

Another interesting insight was found regarding synaesthesia. When asked at the start of the study, 75% (15) stated that they did not have synaesthesia, with 20% (4) stating *"I Don't Know"*, and 5% (1) citing *"Yes"*. While no participants in this study stated that they had musical pitch-colour synaesthesia, there were two participants that recorded scores close to the level of a synaesthete. One participant scored a 1.81 on the first test (Desktop), then scored a 1.10 on the second test (VR), and another participant scored similar scores, with a 1.35 on the first test (VR) and a 1.40 on the second test (Desktop). The proposed synaesthesia score level is set at < 1.0, however, research has shown that this can be conservative and it has been proposed that a score can be set marginally higher to be classed as a synaesthete (Carmichael et al., 2015). These scores would have to be tested and compared across multiple times period to seek a truly consistent score; however, as previous stated, some research posits synaesthesia should be viewed on a spectrum rather than a binary condition (Campen, 2010; O'Toole et al., 2019).

These additional observations give some more general insights into pitch-colour associations and their relationship with synaesthesia beyond VR focus. It highlights that there is a lot to consider when exploring perceptive aspects between our senses, such as past experience, and how different each of us are in how we process associations like those between pitch and colour. There may be differences between people in how they form pitch-colour associations. Some individuals may have more generalised cross-modal associations, while others might have personal associations similar to those of synaesthetes. Can VR help in understanding these connections?

Bowman and McMahan (2007) highlight that VR can become a valuable tool for training and testing in many areas, as long as it can provide a level of immersion that is realistic, producing a sense of presence. With a deeper exploration of the immersive features that VR can offer, this study has shown that there is potential for pitch-colour cross-modal associations to benefit from this technology.

4.11 Final Conclusion

This study examined the potential impact of testing environments, namely a VR head-mounted display (HMD) and a traditional desktop setup, on the performance outcomes of a musical pitch-colour association test. A balanced A/B testing approach involving 20 participants was employed for the investigation. Our primary research question, *"Does environment (Desktop vs VR HMD) significantly affect how participants perform in a musical pitch-colour association test?"* found its answer leaning towards the null hypothesis. The outcomes revealed no significant difference in the performance of participants under the VR and desktop conditions.

Pitch-brightness cross-modal associations were found to be present in the VR environment, in a similar way to the desktop setting. This supports the findings of Malpica et al. (2020), which demonstrated that auditory-visual cross-modal associations are present in VR settings. The current study extends this work by showing that pitch-colour cross-modality is also present—an aspect not previously investigated.

Given that even a low-immersion VR environment offered participants benefits in terms of immersion, focus, engagement, and interaction, shifting the focus toward more immersive features and the full interactive potential of 3D VR environments appears to be a natural next step for this research. This study has been an informative first step into VR's potential for understanding the relationship between pitch and colour for use in consistency tests and also for future design knowledge on how to leverage cross-modal associations for a variety of goals, such as learning. VR's transformative potential has been recognised in related research domains (see Section 4.2.3), and it offers a novel approach for advancing cross-modal research by enabling new forms of sensory integration and interaction.

4.12 Chapter Reflection

This chapter presented an extended version of a published study that explored the use of a VR head-mounted display (HMD) compared to a standard desktop when completing a pitch-colour association test, as part of the online `Synaesthesia Battery`. The study aimed to address the thesis sub-question outlined in Section 1.3, *"Is VR a viable platform for conducting pitch-colour association tests compared to desktop environments?"*

The insights gained from this study contribute to answering this question and begin to build the knowledge required to address the central thesis question: *"How can cross-modal associations of pitch and colour be enhanced through virtual reality?"*

As no prior studies were found that explored pitch-colour cross-modality in VR settings, it was first necessary to determine whether participants perform differently in a VR association test compared to a traditional desktop environment. Additionally, it was important to explore how the use of VR, in the absence of immersive design features such as realistic 3D environments, locomotion, interactive elements, or spatial audio, influences participant scores.

The results validated the hypothesis that no significant differences in test scores exist between the desktop and VR HMD settings. They also showed that pitch-brightness cross-modal associations are present in VR, which is a novel finding. The findings suggest that VR is a suitable environment to explore pitch-colour cross-

modal associations and that the mere use of a VR headset, without the inclusion of immersive features, does not influence participants' pitch-colour associations. This outcome is important as it provides a foundation for the subsequent studies that explore the potential of VR's immersive capabilities to strengthen cross-modal associations.

The study also revealed a user preference for taking the test in VR, which may be attributed to the novelty of the technology or the more intuitive nature of the controls. While the VR environment in this study was intentionally minimal—featuring only the test webpage surrounded by a dark virtual room—this nonetheless highlights the broader potential of VR to create varied and focused environments that may help users concentrate on tasks while also enhancing engagement through spatial immersion.

The next study, presented in Chapter 5, builds directly on these findings and investigates the question: *"What role do immersive features play in supporting pitch-colour cross-modal associations in VR learning environments?"* While this first study confirmed that a low-immersion VR setting does not affect test outcomes, it set the stage for investigating how immersive elements of VR might influence cross-modal associations in future studies. Therefore, the following study shifts focus to a more user-centred approach, employing a mixed-methods design to explore how specific immersive features—such as environment design, spatial audio, interactivity, and user control—may contribute to user engagement and cross-modal understanding.

This progression moves the research beyond desktop-based testing and towards more embodied, perceptual experiences in VR. Such an approach aligns with the core aims of Human-Computer Interaction (HCI) research: to develop human-centric systems that effectively leverage multiple sensory modalities to enhance interaction and learning.

Chapter 5

Investigating Musical Pitch-Colour Associations in Immersive Virtual Reality Environments

This chapter was published as: P. O'Toole, M. Mancini, I. Pitt and Laura Maye. Investigating Musical Pitch-Colour Associations in Immersive Virtual Reality Environments. *Virtual Reality Springer Journal 29, 88, 2025, pp. 1-23.* DOI: 10.1007/s10055-025-01163-8.

The following study explores the sub-question: *What role do immersive features play in supporting pitch-colour cross-modal associations in VR learning environments?*, that is presented in Section 1.3.

Abstract: *Virtual Reality (VR) offers a unique platform to explore auditory-visual cross-modal associations, such as mappings between musical pitch and visual colour, through immersive and interactive environments. This study investigates how Virtual Environment (VE) design and immersive VR features, such as environment setting, auditory and visual elements, locomotion and interactions, can support users in engaging in pitch-colour cross-modal tasks. Thirty-eight participants completed the tasks in two distinct VR environments: "forest" and "space". An A/B experimental design was used, collecting qualitative feedback, usability ratings, and measures of presence, cognitive load, and simulator sickness. Additionally, completion times and pitch-colour/size selections were recorded for each environment. Results revealed that the "forest" environment induced higher simulator sickness compared to the baseline scores (significantly) and the "space" environment. The template analysis highlighted the elements participants found most engaging and immersive, with the "space" environment being more effective at helping participants focus on the task. Preferences varied, but many participants valued 3D audio presentation and visual feedback in the "forest" and appreciated the ability to control locomotion style over a fixed locomotion style. This study informs the design of virtual learning platforms that enhance pitch-colour cross-modal associations, leveraging synaesthetic principles for more engaging and effective VR learning experiences. Beyond these specific applications, the results highlight the potential of XR technology to address broader challenges in tailoring immersive experiences to*

diverse real-world needs and user preferences.

5.1 Introduction

A Virtual Environment (VE) is an immersive, interactive space that simulates real or imagined settings digitally (Brooks, 1999). Immersive VEs have shown a lot of potential in not only providing a research space for exploring auditory-visual multi-modality, but also cross-modal associations between the two senses. Multi-modal interactions are when the different modalities interact with each other (Jacko and Sears, 2003) (e.g., a virtual reality system where users employ both hand gestures and voice commands to manipulate objects), whereas cross-modality is when the different modalities *influence* each other (Guo and Jiang, 2023) (e.g., higher pitched sound are perceived higher when placed in a higher spatial location). Cross-modal associations have been observed between various sensory features across the five senses (Spence, 2011); however, in the context of this study, the term cross-modal specifically refers to associations between pitch in the auditory modality and colour in the visual modality.

Research has demonstrated that cross-modal associations between auditory and visual stimuli occur across multiple features, including pitch, key, timbre, loudness, and spatial or temporal properties in sound, as well as colour, size, shape, and spatial or temporal aspects in vision (Spence, 2011; Palmer et al., 2013; Parise and Spence, 2012). This study focused on pitch-colour cross-modal associations, while also exploring pitch and size associations as a secondary association task.

Pitch has been consistently identified as a salient feature in cross-modal perception. Higher-pitched sounds are strongly associated with lighter colours and smaller sized objects, while lower-pitched sounds correspond to darker hues and larger sized objects (Parise and Spence, 2012; Curwen, 2018; Evans and Treisman, 2010; Gallace and Spence, 2006). This relationship extends to synaesthesia, a phenomenon in which the stimulation of one sensory modality involuntarily triggers perceptions in another (Cytowic, 2003). Chromaesthesia (sound-colour synaesthesia) occurs when auditory stimuli, such as musical notes, consistently evoke specific colours in an individual's perception (Ward et al., 2006).

Extended Reality (XR) technologies encompass Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), each offering different levels of interaction between the digital and physical world—VR provides full immersion in a computer-generated environment, AR overlays digital content onto the real world while maintaining a direct view of it, and MR extends AR by enabling real-time interaction between virtual and physical elements (Vasarainen et al., 2021). XR applications have demonstrated significant benefits in multi-modal design across various domains, including education (Dede, 2009), healthcare (Poguntke et al., 2019), and music (Trujano et al., 2018)—the focus of this study. While AR and MR are useful for blending digital and physical content, this study focuses on VR to explore fully immersive environments. By transporting users into distinct virtual settings, such as a forest or outer space, VR provides a controlled means of examining how environmental design influences experience and engagement—an effect harder to achieve when the real world remains visible in AR or MR.

This research investigates how immersive VR experiences can support pitch-colour and pitch-size cross-modal perception within the context of a simple musical

scale. Specifically, it examines the influence of VE design—encompassing virtual setting, locomotion and interaction controls, and auditory-visual presentation—on auditory-visual perception. With a long-term vision of developing an interactive music-learning experience, this study aims to uncover how VR environments can refine cross-modal design and optimise multi-sensory learning. The following research questions guide this investigation, exploring the potential of immersive VR applications in enhancing pitch-colour associations.

- RQ1: How do the two VR environments impact user experience and performance in the pitch-colour association task?
- RQ2: How do changes between the two environments affect participants' experiences?
- RQ3: What improvements would make the system more effective for the pitch-colour association task?

The two environments investigated in this study include a natural, visually rich "forest" environment, and an otherworldly, minimalist "space" environment. Exploring these two contrasting environments provides insights into optimising virtual environment design to harness auditory-visual cross-modality for music-learning platforms, while also offering potential benefits for other educational contexts.

The reason this study focuses on the "forest" and "space" VEs is that they both provide qualities that can optimise user's engagement when conducting cognitive tasks, such as pitch-colour cross-modal association tasks. As this paper will highlight below and in Sections 5.2.4 and 5.4.4, these VEs can serve as immersive platforms that minimise cognitive stress, promote relaxation and engagement, and enhance presence. By exploring both a visually rich natural environment and a minimalist setting, this study aims to examine how VE design—including auditory-visual presentation, locomotion, and interaction—shapes cross-modal associations between musical pitch in the auditory sense and colour and size in the visual sense. The combination of these elements influences immersion, cognitive load, and engagement, ultimately affecting how users perceive and experience pitch-colour and pitch-size relationships in VR.

Studies have shown that full immersion in VR is not always needed or wanted to provide an immersive experience to the user (Bowman and McMahan, 2007; Grimshaw-Aagaard and Walther-Hansen, 2024). The level of immersion in a VR environment can significantly impact user experience, cognitive load, and learning effectiveness (Makransky et al., 2019). While high immersion can enhance presence, it may also introduce distractions or increase cognitive demands, depending on the design of the environment. Minimalist environments, such as a "space" environment, have been shown to reduce cognitive load, minimise distractions, and enhance memory and spatial accuracy, making them effective for focused learning tasks (Rettinger et al., 2024; Gkoumas and Shimi, 2022). Conversely, natural environments, such as forests, have been found to promote relaxation, reduce anxiety, and stimulate creativity (Yu et al., 2018; Lee et al., 2022; Mattila et al., 2020), which may positively influence engagement in musical learning tasks.

Beyond the auditory-visual presentation, locomotion and interaction also play a role in creating immersive experiences, shaping presence, engagement, and cognitive

load in VR. Research suggests that active movement and hands-on interaction enhance experiential learning by boosting motivation and retention (Mantovani, 2001). This study examines how varying locomotion and interaction styles across the two environments influence immersion and the formation of cross-modal associations between musical pitch and colour.

To support this investigation, the following sections are presented. The Related Work section reviews prior studies on auditory-visual associations, multimodal interactions in XR, immersion and presence in VR and the impact of virtual environments on user experience. The Methodology section describes the application setup and environment design, captured data, experimental design, as well as participant selection, data collection, and analysis methods. The Results section presents the key findings from both qualitative and quantitative measures, followed by the Discussion, which interprets these results in the context of existing research and highlights implications for VR-based cross-modal learning between the auditory and visual senses. The Limitations and Future Work section suggests directions for improvements and future research, with the Conclusion summarises the key insights.

5.2 Related Work

5.2.1 Auditory-Visual Associations

In Section 5.1, cross-modal associations, synaesthesia and the close relationship between the two topics are explained. With regard to the auditory and visual modalities, past research has presented cross-modal associations between many different features, such as, musical pitch-colour, musical key-brightness, timbre-saturation, loudness-brightness and pitch-size (Curwen, 2018; Ward et al., 2006; Palmer et al., 2013; Spence, 2011). These mappings include higher pitches being associated with brighter colours and smaller sizes, while major keys and increased loudness correspond to increased brightness. Conversely, lower pitches, a minor key, and decreased loudness are associated with the opposite characteristics (Curwen, 2018; Ward et al., 2006; Palmer et al., 2013; Spence, 2011).

Cross-modal associations between the auditory and visual senses share certain commonalities with sound-colour synaesthesia with both groups appearing to employ similar pitch mapping with regard to pitch-lightness, however the difference is automaticity, consistency and specificity of colour selections (Ward et al., 2006).

Several consistency tests assess synaesthesia, including the `Synaesthesia Battery` (Eagleman et al., 2007), `Test of Genuineness (TOG)` (Asher et al., 2006), and `Netherlands Colour Synaesthesia (NeCoSyn)` (Campen and Froger, 2003). These tests evaluate sound-colour synaesthesia, with the `Synaesthesia Battery` and `TOG` also assessing other forms, such as grapheme-colour and spatial-sequence synaesthesia. Studies using these tests with both synaesthetes and non-synaesthetes suggest that cross-modal associations exist on a spectrum rather than fitting a binary classification (Campen, 2010). This raises the question: rather than "synaesthete" or "non-synaesthete", to what extent is one synaesthetic? (Campen, 2010).

With the unique environment that VR can offer, in terms of immersion, presence and engagement, could it be a viable tool for exploring cross-modal associations between the auditory and visual senses? Merleau-Ponty (2002, see p.266) states: *"synaesthetic perception is the rule and we are unaware of it only because scientific*

knowledge shifts the centre of gravity of experience, so that we have unlearnt how to see, hear, and generally speaking, feel, in order to deduce, from our bodily organisation and from the world as the physicist conceives it, what we are to see, hear, and feel". Here, he argues that our natural synaesthetic perception is overshadowed by scientific methods that separate sensory experiences. One theory suggests that we are all born with synaesthesia-like neural hardware, where our senses are mixed and have not yet been organised into separate modalities, and most of us lose this during the natural course of development (Ward et al., 2006). Potentially, VR could harness this "unlearnt" synaesthetic perception by immersing users in environments where sensory inputs are seamlessly integrated, enhancing our ability to experience and interact with blended sensory information.

Synaesthesia has already inspired research in more artistic areas, like enhancing music perception (Rudenko and Cabral, 2019), teaching creativity (Murgia, 2015), and multi-modal art (Shirokova, 2021). More recently, VR technology has been used in synaesthesia research (Silva Bastos et al., 2018; Bordegoni et al., 2023; Rudenko et al., 2022; Weinel, 2021).

As the technology improves, VR will become a prominent tool for research, and the exploration of cross-modality and how it can potentially be used to enhance our perception in music, education, art and gaming. By exploring the integration of auditory and visual elements such as pitch-colour and pitch-size associations, it contributes to the understanding of how XR tools can be tailored to specific user experiences.

5.2.2 Multi-Modal Experiences in Virtual and Extended Reality

Extended Reality (XR) technologies—including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR)—enable multi-modal interactions, making them valuable tools for education. Research into AR and MR for piano instruction has demonstrated their ability to enhance musical skill development through interactive and engaging learning methods (Trujano et al., 2018; Molero et al., 2021). Beyond music, XR has been shown to support learning in social and natural sciences (Pellas et al., 2020) and accessibility education (Lorenzo et al., 2022; Rudenko and Haahr, 2022).

Despite the multi-modal potential of VR, much of its research and development has historically been visually driven, even though auditory stimuli are processed more rapidly and can be just as informative (Kailas and Tiwari, 2021). However, recent studies have increasingly explored audio-first VR experiences, deepening our understanding of VR's role as a multi-modal learning tool (Kailas and Tiwari, 2021; Gröhn et al., 2005; Hirway et al., 2022).

Multi-modal learning has been shown to enhance musical ability, memory, and improvisation skills. For example, an experimental study comparing traditional vs. multi-modal music instruction found that the multi-modal group outperformed the control group, particularly in rhythm, musicality, and memory, with scores exceeding 85.5% (Ou, 2023). Additionally, active auditory engagement has been linked to faster and more robust neural processing of speech (Kraus et al., 2014).

Research has also demonstrated that sound can influence visual perception (Malpica et al., 2020), reinforcing the role of cross-modal interactions in immersive

settings. More recently, VR has become a key tool for investigating cross-modal associations, particularly in understanding how sensory modalities interact in immersive environments (Malpica et al., 2020; Chalmers et al., 2009; Grimshaw-Aagaard and Walther-Hansen, 2024).

With the growing interest in cross-modal perception in VR, researchers have proposed the exploration of synaesthetic-inspired design to enhance immersion and user engagement (Grimshaw-Aagaard and Walther-Hansen, 2024; Wright and Kefalidou, 2021). While traditional VR research often focuses on expected user interactions, recent studies have explored "atypical" interactions, particularly within a synaesthetic design framework (Wright and Kefalidou, 2021).

Since this study examines musical pitch-colour and pitch-size associations, designing "atypical" yet musically familiar interactions may enhance user presence, immersion and engagement. According to Curwen (2018), music-colour synaesthesia is shaped not only by internal cognitive processes but also by real-time interactions between individuals and their environment. This supports the potential of cross-modal VR design to aid users in performing auditory-visual tasks by fostering immersive and engaging experiences.

5.2.3 Immersion, Presence, & Interaction in VR

Immersion, presence and interaction are critical in shaping user experience in VR, particularly when engaging in cross-modal tasks like pitch-colour associations. Understanding how users perceive and interact within virtual environments is essential for designing effective multi-sensory experiences that enhance engagement, cognitive processing, and task performance.

Immersion is an objective measure of a VE's technological fidelity, including sensory richness and responsiveness, whereas presence is the subjective feeling of "being there" (Bowman and McMahan, 2007). While higher pixel density and frame rates enhance immersion, presence also depends on cognitive and perceptual factors beyond hardware alone. Çamcı and Hamilton (2020) notes, immersion and presence are not purely technological but require interdisciplinary research in psychology, perception, and design. This underscores the need to consider both technical and human-centred aspects in VR development.

Recent studies highlight how cross-modal design and spatial audio contribute to enhanced immersion (Silva Bastos et al., 2018; Potter et al., 2022). Research suggests that high-quality audio, even with lower visual resolution, can provide equivalent or greater immersion compared to high-resolution visual scenes alone (Potter et al., 2022). Similarly, Grimshaw-Aagaard and Walther-Hansen (2024) propose "reduced realities", where strategic cross-modal design enhances presence without overloading the user. These findings suggest that well-designed multi-modal VR experiences can achieve high levels of immersion without requiring intensive visual fidelity.

Locomotion is essential for navigating VR environments and includes various methods such as physical walking, walking in place, controller-based movement, teleportation, leaning, and flying (Boletsis and Cedergren, 2019; Wilson et al., 2016). In commercial VR, controller-based locomotion and teleportation are widely used due to their suitability for small spaces and minimal physical effort. However, different locomotion techniques impact immersion and user comfort in distinct ways.

A study by Boletsis and Cedergren (2019) found that walking in place provided

high immersion but caused physical discomfort, whereas controller-based locomotion was user-friendly. Teleportation, while effective, was found to disrupt immersion due to visual jumps. Similarly, Frommel et al. (2017) compared free teleportation, fixed teleportation, touch-pad movement, and guided locomotion in a VR zoo environment. The study found that free teleportation was preferred for presence and enjoyment, whereas touch-pad locomotion led to higher simulator sickness scores, and guided locomotion was the least preferred overall.

In addition to locomotion, interaction methods significantly affect immersion and usability. VR interactions typically involve direct interactions, such as hand tracking, gaze, and pointers; physical controls, including joysticks, steering wheels, and buttons; and virtual controls, which rely on in-world User Interface (UI) elements (Mine, 1995). Physical controls tend to enhance presence and precision, whereas virtual controls lack haptic feedback but allow for intuitive in-world interaction. Object interaction methods in VR can be categorised into "at a distance" (ray interactor) and "local" (direct interactor) approaches. Local interactions are preferred for precision tasks, while distance-based interactions are beneficial in environments requiring broad navigation.

Finding balance between the auditory, visual, locomotive and interaction features is the challenge when designing immersive engaging virtual environments. Previous research has shown that full immersion is not always needed and that the amount of immersion can be application-dependant (Bowman and McMahan, 2007). The exploration into enhancing pitch-colour perception using cross-modal associations in VR settings is a new research area, and gaining insights into the different VR features and the virtual environment the users are set in will provide a starting point for immersive platforms that harness cross-modal design in different applications.

5.2.4 Virtual Environments & User Experience

The design of virtual environments (VEs) significantly influences immersion, presence, user engagement, performance and cognitive load. As stated in Section 5.2.3, full immersion is not always needed and clever design can provide immersion without the need for high-quality technical solutions. This section compares and contrasts prior research on visually complex and minimalist environment design in VEs, highlighting the different effects each can have on users.

When looking at visually complex versus more minimalist VE design, both approaches have positive and negative consequences. Environments that present realistic spaces with high levels of immersion can enhance presence but may also increase cognitive demands, and affect performance in learning-based tasks (Makransky et al., 2019; Schrader and Bastiaens, 2012). Immersive nature-inspired virtual environments, especially forests, have been shown to have several positive psychological effects for users. It has been found that virtual forests can reduce stress (Valtchanov et al., 2010; Wang et al., 2019), have a positive effect on emotions or mood (Mattila et al., 2020; Yu et al., 2018; Lee et al., 2022; Valtchanov et al., 2010), increase vigour (Yu et al., 2018), provide cognitive recovery (Lee et al., 2022) and improve creativity (Nanjappan et al., 2023) for users. Furthermore, Gómez Sirvent et al. (2023) found that exposure to natural elements, such as light and window views, in VR provides a more positive evaluation of the environment, compared to environments with an absence of daylight and no outside views.

In contrast, minimalist environments are designed to reduce cognitive distractions and enhance task efficiency. While increased presence is usually tided to higher immersion (Makransky et al., 2019), some research has explored other ways of increasing presence without overloading the senses (Grimshaw-Aagaard and Walther-Hansen, 2024). Research suggests that lower visual complexity is associated with reduced cognitive load, making minimalist environments particularly effective for learning and memory retention (Rettinger et al., 2024). Gkoumas and Shimi (2022) found that simpler environments improve spatial accuracy by minimising unnecessary distractions, allowing users to focus more on the task at hand.

As Potter et al. (2022) highlight, visual complexity and realistic graphics are not always necessary for creating immersive experiences. High-quality experiences can be achieved through the effective integration of multiple sensory modalities. Exploring different virtual settings, locomotion types, and audio-visual presentations will help identify the key features that enhance immersion, presence, and engagement while minimising negative effects such as cognitive load and simulator sickness. These insights will contribute to the development of future cross-modal learning experiences in XR.

5.3 Methodology

This study compares two VEs, both with distinct VR features described in the subsequent Section 5.3.1. Participants will engage in a pitch-colour and a secondary pitch-size association task involving musical notes in the C Major scale. The two environments under investigation are a visually rich natural "forest" and a minimalist, otherworldly "space" environment. These environments differ in terms of visual setting, audio presentation, auditory-visual feedback, and locomotion types, as described in the subsequent Section 5.3.1.

The following subsections will detail the environment setup & design, captured data, experiment design, participant selection, data collection methods, and data analysis procedures.



Figure 5.1: Left: Participant in the soundproofed studio engaged in the VR application, and Right: Image of the "forest" environment with the fog that provides an extra colour feature

5.3.1 Environment Setup & Design

The different environments under investigation in this study were designed using the `Unity Game Engine`, and tested to run on an `Oculus Quest 2` with controllers. This study requires users to conduct cross-modal tasks, where they associate musical pitch with colour and size, while experiencing two different environment setups. Previous research has found cross-modal correspondence between auditory pitch-lightness (Marks, 1987; Ward et al., 2006), pitch-brightness (Marks, 1987; Spence, 2011) and pitch-size (Evans and Treisman, 2010; Gallace and Spence, 2006; Spence, 2011).

Users will navigate two environments and interact with musical orbs that play notes in the C Major scale, ranging from C4 (261.6 Hz) to C5 (523.3 Hz). Each note of the scale will be played twice per environment, with the order randomised (16 notes x 2 environments). Additionally, users will adjust the orb's size to match their perception of the musical pitch by moving their hands apart to increase the size or together to decrease it. The study examines how pitch-colour and pitch-size associations are performed in two different environments, aiming to identify the optimal setting for conducting cross-modal tasks focusing on music exercises.

As highlighted in Section 5.2.4, different VE settings come with positive and negative effects that can impact the presence felt and the overall user experience. The two environments chosen for this study were a visually rich "forest" and a minimalist "space" setting. With these two contrasting VEs, which both provide characteristics that can help foster comfort, presence and engagement in cognitive tasks, an exploration into how participants feel, engage and perform the pitch-colour and pitch-size tasks is conducted.

This study presents four different virtual scenes to users, following this sequence: a tutorial scene, followed by Environment A ("forest" or "space"), then a questionnaire scene, followed by Environment B ("forest" or "space"), and concluding with a final questionnaire scene (see Figure 5.3). The following sections will describe the design of each of the VR scenes experienced in this study, starting with the "forest", the "space" scene, the tutorial and finally the questionnaire scene.

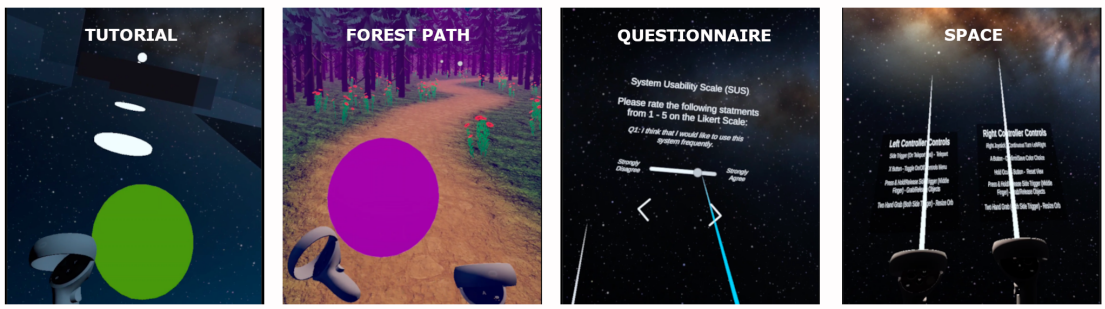


Figure 5.2: The four environments experienced in virtual reality: Tutorial scene (far left), Forest Path (centre left), Questionnaire scene (centre right), and Space scene (far right).

Forest

The "forest" was chosen for this study because of its familiarity and its positive impact on people. As highlighted in Section 5.2.4, studies show that virtual forests relieve stress, enhance creativity, and decrease negative emotions, which will aim to

provide a calming setting for participants to conduct the pitch-colour and pitch-size association tasks.

The "forest" provides a more realistic setting to conduct the cross-modal association tasks, it was designed to enhance immersion and realism while minimise negative aspects such as cognitive load as much as possible. In this setting, users engage in pitch-colour and pitch-size association tasks while navigating across a "forest" environment bordered by grass, trees, flowers, and rocks, with fog drifting from the trees and mountains in the background.

The environment uses free movement with the game-pad for locomotion and snap-turning in 30-degree increments for turning. While free movement is common in VR games, it can cause more motion sickness compared to other methods (Cherni et al., 2020). Snap-turning, which turns the avatar in fixed increments, has been shown to reduce motion sickness compared to other turning controllers in VR (Sargunam and Ragan, 2018). This environment uses `Unity`'s 3D spatial audio effects for environmental sounds like trees, birds, wind, and footsteps to enhance presence and the musical notes triggered by interacting with the orbs.

Visually, the orb changes colour based on the controller movement on the y-axis of the hand not interacting with the orb. Additionally, fog colour matches the orb's colour to provide a synaesthesia-like experience when auditory stimuli are heard (see Figure 5.1).

With Research showing that VR forest environments decrease negative emotions, provide cognitive recovery and stress relief (Yu et al., 2018; Lee et al., 2022; Mattila et al., 2020; Wang et al., 2019), this environment should provide a relaxing and immersive experience to the participants while conducting the cross-modal task.

Space

The "space" VE design adopts a more minimalist approach with an otherworldly experience. This virtual environment features the `Milky Way` galaxy as its backdrop, using fixed teleportation and continuous turning to navigate. Fixed teleportation typically scores low on `SSQ` tests, with free teleportation scoring the lowest (Frommel et al., 2017). Users start by moving to a white teleport pad, which relocates randomly after each task, guided by a 3D audio locator. Once the user is teleported to the new pad, a white orb appears, allowing interaction in the same manner as in the "forest path".

The two key differences in the interactions with the musical orbs in the "space" environment, compared to the "forest", are that the audio design is not spatialised, and there is no additional visual effect, such as the fog effect present in the other environment. After completing all sixteen locations, a message indicates the next scene is loading.

Research indicates that simpler environments reduce distractions, improving memory and location accuracy (Gkoumas and Shimi, 2022). A study on "reduced realities" supports a "less is more" approach, enhancing presence and attention while reducing stress through cross-modal design (Grimshaw-Aagaard and Walther-Hansen, 2024).

Tutorial

Following an internal usability assessment, where the VR application was tested for bugs and overall flow by the research team, it was determined that a more informative tutorial could potentially reduce cognitive overload before participants engaged with the main study. To address this, a minimal glass-enclosed environment was created, allowing users to practice the locomotion and interactions required for the pitch-colour/size tasks.

Mayer (2023) identifies three effective techniques for managing cognitive load: narrative instructions, segmenting information, and signalling. In the tutorial, verbal instructions are delivered as a narrative, with the information needed segmented to allow users to focus on one task at a time. This approach, along with a four-minute timer on the right controller paired with 3D audio cues, collectively seeks to reduces cognitive load by enhancing comprehension, providing clear structure, and pacing the practice period.

Users can then freely explore the tutorial space, which functions as a "sandbox" (a testing area in a computer application) environment that aids in learning game mechanics and reducing cognitive load before the main environments are entered (Jerald, 2015). When ready, a yellow teleport pad guides them to a general questionnaire that collects basic information and offers further practice with `UI` controls needed for the questionnaire scene.

Questionnaire Area

After experiencing each environment, participants enter a dark 3D space to answer questions related to the `System Usability Scale (SUS)`, `iGroup Presence Questionnaire (IPQ)`, and the `Cognitive Load Measurement (CLM)` using a UI slider (see Figure 5.2). The `SUS` has ten items on a five-point Likert scale, the `IPQ` has fourteen items on a seven-point scale, and the `CLM` uses a nine-point scale with one item. Audio instructions outline the number of questions and scale points for each questionnaire. A study comparing VR-based and non-VR-based presence questionnaires found minimal differences, except for reduced testing time and disorientation in VR (Schwind et al., 2019).

5.3.2 Captured Data

Measuring the usability and effectiveness of VR systems requires specialised tools due to the unique nature of immersive environments. While the `SUS` is a widely-used tool for assessing system usability, it is insufficient on its own for VR applications (Brooke, 1996; Lewis, 2018). Additional assessment scales help give a complete picture of an application, accounting for presence, cognitive load, and simulator sickness.

Extensive research has led to the development of various tests designed to measure presence (Witmer and Singer, 1998; Usoh et al., 2000; Schwind et al., 2019; Schubert et al., 2001), cognitive load (Paas, 1992; Leppink et al., 2013; Hart and Staveland, 1988), and simulator sickness (Kennedy et al., 1993; Kim et al., 2018; Stone, 2017). While these tests have been broadly applied, recent studies have increasingly focused on their application within virtual reality environments.

One factor to consider when measuring simulator sickness with a questionnaire like the Simulator Sickness Questionnaire (SSQ), is how the participant is feeling before putting on the VR HMD. One study found that using a baseline score of zero is not representative of users' well-being before entering a VR environment. Their study presents evidence that a zero baseline may be inherently incorrect, as users may already present symptoms mentioned on the SSQ before entering the VR environment (Brown et al., 2022).

For this study, the IPQ, the CLM and the SSQ were chosen as well as the SUS to gain in-depth understanding of the important aspects of the VR experience. The selection of standardised assessment scales was guided by the necessity for prompt feedback immediately following the experience, as well as the utilisation of validated scales recognised and endorsed by the research community.

In addition to the standardised assessment scales, data on pitch-colour and pitch-size associations, as well as completion times for each environment and interaction, were recorded.

5.3.3 Experimental Design

This study seeks to understand how participants engage in two distinct VEs while performing a pitch-colour and pitch-size association tasks. The use of quantitative and qualitative data helped to understand how users engaged with the two environments. SUS, IPQ, CLM, and the SSQ were used to capture data to understand the participants' experience of the application in terms of satisfaction, presence, cognitive load and amount of simulator sickness. Also, colour choices were recorded for each auditory object as well as completion time data for each environment and interaction.

In contrast, a post-study questionnaire was used to capture a more in-depth understanding of the participants' thoughts on the VR experience. An A/B test design was adopted, with all users experiencing both environments, with half the participants starting with the "forest" and the other half starting in the "space" environment.

The study was conducted in a quiet lab environment (see the left image of Figure 5.1). The hardware used for the study was an Oculus Quest 2 HMD with two controllers and a pair of Sony MDR-CD580 headphones. Participants could adjust the volume on the Quest 2 HMD to a comfortable level.

Participants completed the SSQ at three stages: before starting the study ("baseline"), after experiencing the first environment ("forest" or "space"), and after experiencing the second environment ("forest" or "space"), all outside of VR. Providing a "baseline" score for the SSQ was considered as a more reliable way to effectively measure the symptoms on the SSQ, instead of just providing a zero baseline score (see Section 5.3.2). The SUS, IPQ, and CLM questionnaires were administered within the VR application, right after experiencing the "forest" and "space" environments. A more in-depth description of the VR application and each environment is given in Section 5.3.1. The researcher gave a brief explanation of the application controls and the task involved before the participant put on the HMD and entered the application's tutorial environment.

Both environments included the same pitch-colour and pitch-size association tasks. The participant was instructed to grab each of the 16 white orbs with the

side-trigger on the controller, which would play a musical note and allow the user to manipulate the orb's colour and size. The musical note was played randomly from the 8 notes between C4 and C5 on the C Major scale, and each note was played twice (16 interactions). Once the participants had saved all the pitch-colour and pitch-size associations, they were asked to answer the `SUS`, `IPQ`, and `CLM` questionnaires. A five-minute break was taken after the questionnaires were answered and the second `SSQ` was taken. This process (minus the tutorial) was repeated with the other environment, with the final `SSQ` being filled out before the post-test questionnaire was conducted. The experimental procedure, shown in Figure 5.3, was followed by all participants, lasting roughly 50 minutes, with the mean time of 38 minutes and 11 seconds from first entering VR to exiting the VR application.

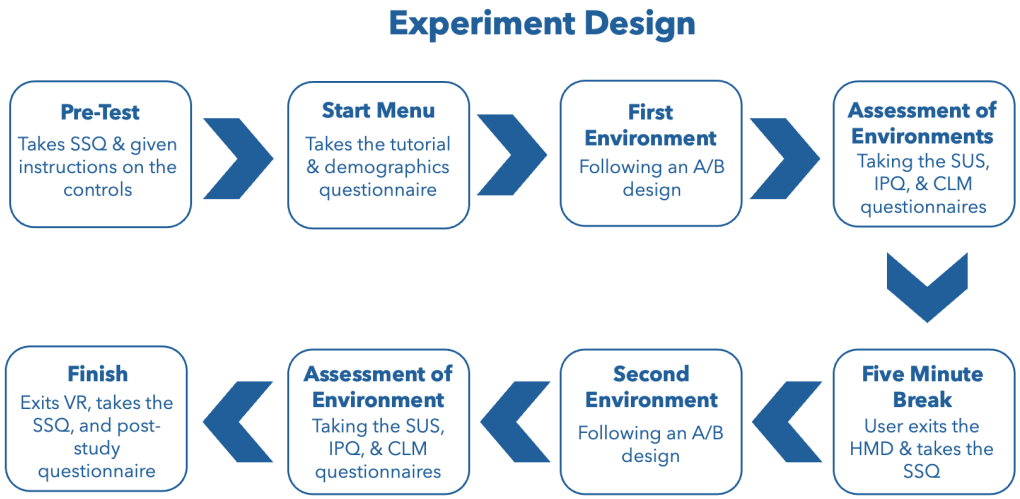


Figure 5.3: This figure shows the study design for Study Two, outlining the sequence of steps each participant followed and indicating when the different subjective scales and questionnaires were completed after each environment.

5.3.4 Participants

Thirty-nine participants took part in the study; however, one participant was unable to finish the test due to time constraints. From the remaining 38 participants, 23 were male, 12 were female, 2 identified as non-binary, and 1 participant did not disclose their gender. 47.4% of the participants were aged between 26-35, with 26.3% from 36-45, 13.2% from 18-25, 7.9% from 46-55, and 5.3% from 56-65 age groups.

Twenty-seven participants had little to no VR experience, three had 1-3 years, three had 3-5 years, and five had over 5 years of experience, respectively. Participants also had varying levels of experience playing a musical instrument, with twelve having little to no experience, twelve with 1-3 years, two with 3-5 years, four with 5-10 years, and eight with 10+ years of experience.

Player IDs where incremental from "Player_ID01" to "Player_ID44". The additional five IDs generated beyond the initial number of participants were due to a

bug in the "tutorial" section that required participants to restart the application. No participants reported having any issues with motion sickness before starting the study. The University College Cork (UCC) Ethics Review Board granted ethics approval (Log 2023-199), and the study was conducted in accordance with the Declaration of Helsinki.

5.3.5 Data Collection & Analysis

Most of the data were collected within the VR application and were saved on the headset in a `JavaScript Object Notation (JSON)` file format named by each player's ID. The saved `JSON` file captured general demographic information (age group, gender, VR, and musical experience), start and end timestamps, pitch-colour/size association data for each object, and responses to the `SUS`, `IPQ`, and `CLM` scales. The `SSQ` questions were filled out with pen and paper before the study started and after the first and second environments. The results were scored and transferred to an `Excel` file for analysis. `Microsoft Forms` was used to capture the post-study questionnaire data. All data were associated with the player's ID, with participants' anonymity maintained across the different data collection methods.

For the data analysis, `Python` was used to read and structure the data from the thirty-eight `JSON` files into a readable format. All the `SUS`, `IPQ`, and `CLM` scores were calculated and saved along with their sub-scale scores and demographic information to a `CSV` file format. `SPSS` was used to run the `Wilcoxon Signed Rank Test`, and the `ANOVA`, while `R Studio` was used to conduct a crossover robust linear model.

To assess pitch-colour associations, a choice variance score was calculated for each participant in each environment by tallying how often each pitch-colour combination was selected across both repetitions, then computing the variance across all combinations. In this context, a higher variance score reflects more consistent associations (i.e., combinations were selected either 0 or 2 times), while a lower variance indicates more uniform but weaker associations (e.g., combinations selected only once).

Regarding the qualitative data, a template analysis was conducted using `NVivo 12` to code the participants' insights from the post-test questionnaires. The participants' insights were coded by the theme of each question and then by different features that had an impact on that theme. Relevant insights cover the four main features under investigation: environment setting, audio elements (environmental audio and auditory feedback), visual elements (visual setting and visual feedback), and controls.

5.4 Results

The results of this study are presented in four parts: first, a general overview based on participants' feedback on their preferred environment and features; second, the system's usability results from the standardised assessment scales; third, the results from the pitch-colour/size associations and the environment completion times; and finally, the template analysis results from the participants' feedback from the post-study questionnaire (See in Appendix B.5).

Table 5.1: Results from the Wilcoxon Signed Ranked Test, along with each environments mean score percentage on each assessment scale

Assessment Scale	Forest (Mean)	Space (Mean)	Sig
SUS (System Usability Scale)	76.38%	74.87%	0.646
IPQ-GP (General Presence)	71.93%	66.67%	0.346
IPQ-SP (Spatial Presence)	71.41%	71.58%	0.694
IPQ-INV (Involvement)	64.03%	62.71%	0.475
IPQ-REAL (Realism)	34.65%	34.54%	0.858
IPQ-OP (Overall Presence)	58.74%	57.96%	0.994
CLM (Cognitive Load Measure)	50.33%	58.55%	0.086

5.4.1 General Overview

When participants were asked, *"Which environment did you feel more comfortable in?"*, responses were evenly distributed: 36% preferred the "forest", 33% favoured "space", and 31% expressed comfort in both environments. For the question, *"Please indicate your preferred environment for each of the listed features below"*, with features including locomotion, audio, visuals, and surroundings, preferences were as follows: 53.8% favoured "space" for locomotion, 69.2% preferred the "forest" for audio, 53.8% slightly preferred the "forest" for visuals, and 61.5% favoured the "forest" for surroundings.

5.4.2 Standardised Assessment Scales

Table 5.2: Results from the Pairwise Comparison following a repeated Measures ANOVA test on the SSQ subscales. * shows marginal significance and ** shows significant p-values < 0.5.

Environment	Nausea	Oculomotor	Disorientation	Total
Baseline v Forest	.028**	.021**	.022**	.004**
Baseline v Space	.072	.124	.422	.057*
Forest v Space	1.0	.473	.230	.373

First an analysis of the SUS, IPQ, and CLM scores was conducted, the Wilcoxon Signed Rank Test was used as the data was non-parametric (see Table 5.1). No significant differences between the "forest" and "space" were found. The mean percentage scores are presented in the following order ("forest" first, followed by "space"), along with the corresponding p-values. The SUS scores were 76.38% and 74.87%, respectively, with no significant difference between the two ($p = .646$). For general presence (IPQ-GP), the scores were 71.93% and 66.67% ($p = .346$), while spatial presence (IPQ-SP) scores were nearly identical at 71.41% and 71.58% ($p = .694$). Involvement (IPQ-INV) scores were 64.03% and 62.71% ($p = .475$), and realism (IPQ-REAL) was similarly close at 34.65% and 34.54% ($p = .858$). The overall presence scale (IPQ-OP) showed scores of 58.74% and 57.96% ($p = .994$). Finally, cognitive load measurement (CLM) was higher in the "space" environment (58.55%) compared to "forest" (50.33%), with marginal significance ($p = .086$).

To assess whether the sequence of the A/B design and the order in which the environments were experienced affected the scores, a robust linear mixed model was applied. The results indicated no significant effects for sequence and test time across all scales, except for test time in the `SUS` score. Specifically, when the test was taken a second time, a significant increase in the `SUS` score was observed ($t = 3.159$), regardless of the environment is being experienced. This suggests a learning or adaptation effect from having experienced the first environment.

A repeated measures `ANOVA` was conducted to compare the effects of different environments ("baseline", "forest", "space") on the `SSQ` (simulator sickness questionnaire) sub-scales: **Nausea**, **Oculomotor**, **Disorientation**, and **Total** scores (see Table 5.2). A significant effect of environment was observed for **Nausea**, $F(2, 74) = 4.780, p = .011$, with pairwise comparisons indicating higher scores in the "forest" environment compared to "baseline" ($p = .028$). For **Oculomotor**, significant linear ($F(1, 37) = 4.476, p = .041$) and quadratic ($F(1, 37) = 5.864, p = .020$) trends were found, with higher scores in the "forest" environment compared to "baseline" ($p = .021$). A significant effect of environment was also found for **Disorientation**, $F(2, 74) = 5.008, p = .009$, with pairwise comparisons showing higher scores in the "forest" environment compared to "baseline" ($p = .022$). Lastly, for the **Total score**, a significant environmental effect was observed, $F(2, 74) = 7.077, p = .002$, with "forest" yielding higher scores than "baseline" ($p = .004$).

Overall, the "forest" environment elicited significantly higher simulator sickness symptoms compared to "baseline" across all sub-scales, while "space" showed borderline significance compared to "baseline", with no significant differences between "forest" and "space". A robust linear mixed model revealed no significant effects of sequence or test time on `SSQ` scores across all sub-scales.

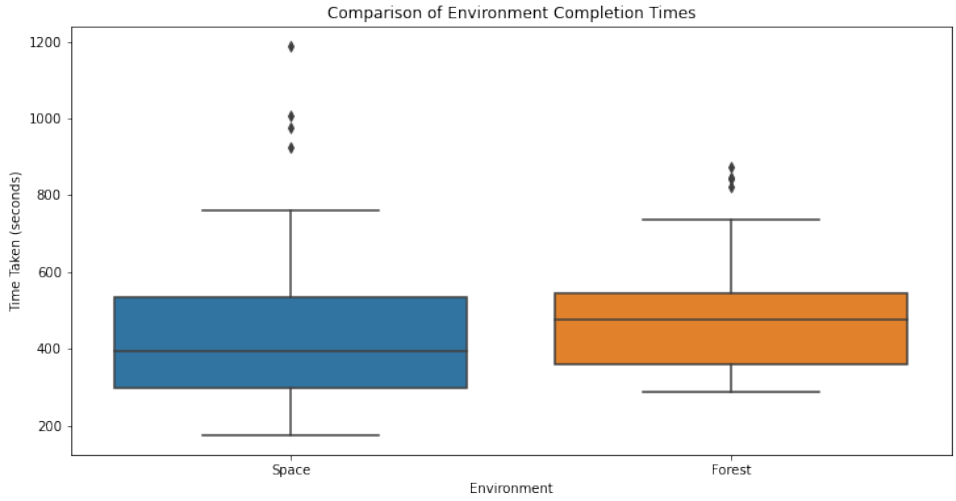


Figure 5.4: Left: the average completion time for participants in the space environment (in seconds); Right: the average completion time for participants in the forest environment (in seconds)

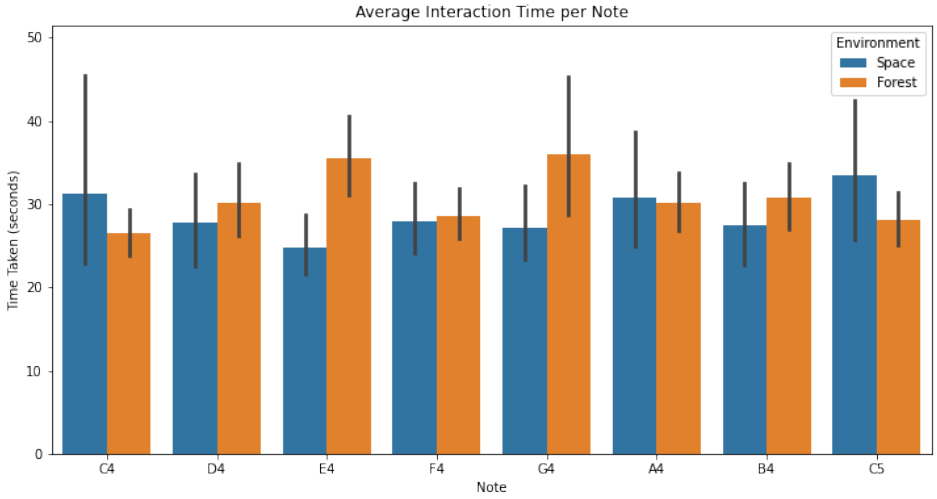


Figure 5.5: Average time taken (in seconds) to make pitch-colour associations for each note, with interactions in the Space environment shown in blue and interactions in the Forest environment shown in orange.

5.4.3 Completion Times & Colour-Pitch/Size Association Analysis

A `Wilcoxon Signed-Rank Test` was performed to compare completion times between the "forest" and "space" environments. The results showed no significant difference ($Z = -1.516, p = 0.130$), indicating that participants took similar amounts of time to complete tasks in both environments. Figure 5.4 presents a box-plot showing the average completion times in the "forest" and "space" environment. While Figure 5.5 highlights the average completion time per note across both environments, with the "space" environments providing lower completion times in five of the eight notes.

A `Wilcoxon Signed-Rank Test` was conducted to compare pitch-colour variance scores between the "forest" and "space" environments. The results showed no significant difference, $Z = -0.78, p = .435$, suggesting that participants exhibited similar levels of colour variance in both environments. In Figure 5.6, a heat map shows the total pitch-colour associations made across the participants for each note and hue comparison.

The secondary task of associating pitch to size was actively engaged with 83% of the time across both environments. Only eleven participants fully engaged with the pitch-size interactions. The heat-map in Figure 5.7 highlights pitch-size associations made by participants over both environments. A high degree of choices across all musical pitches associate with the "medium" size, however higher frequency of lower pitch and larger size associations are found, with the "small" and the "smallest" size rarely chosen.

5.4.4 Template Analysis Results

To investigate the qualitative aspects of the study, a template analysis was conducted on the feedback from participants in the post-study questionnaire (see Appendix

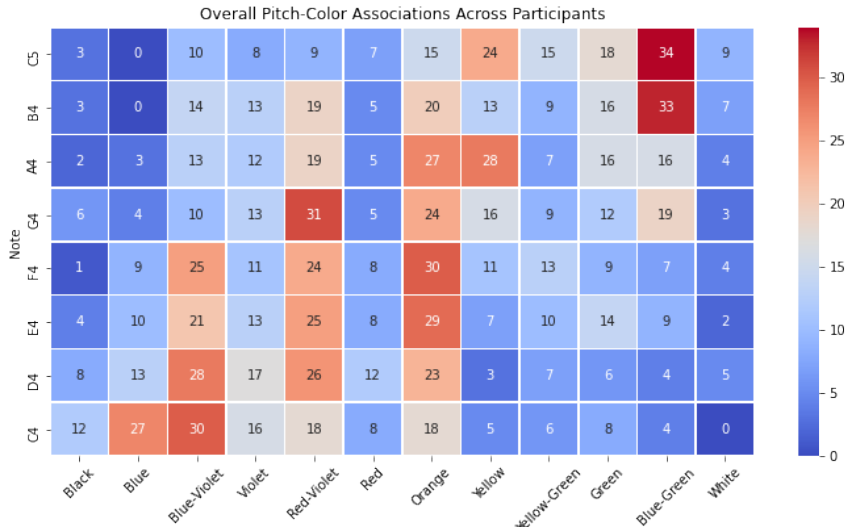


Figure 5.6: This heat map highlights the frequencies of pitch-colour associations participants chose for each note in the C Major Scale. Higher pitch-colour association frequencies as shown as a redder colour, while lower frequencies are shown as darker blues.

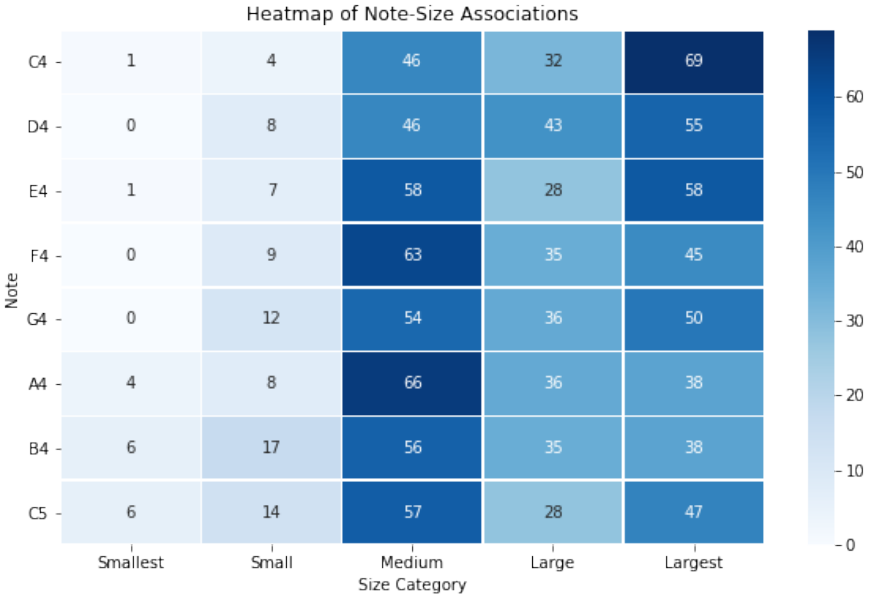


Figure 5.7: This heat map shows which size ranges participants chose for each note in the C Major scale, with size ranges grouped into "smallest", "small", "medium" (default size), "large", and "largest".

B.5). Template Analysis is a form of thematic analysis that emphasizes hierarchical coding, combining structured textual data analysis with flexibility. Researchers develop and refine a coding template from a subset of data, which is then applied to further data, allowing extensive theme development where the richest data relevant to the research question is found (Brooks et al., 2015). `NVivo 12` was used to analyse the data by coding responses into topics such as immersion, ease of use, comfort, and engagement, and then by features like movement, visual feedback, controls, and environmental audio to gain insights into participants' experiences. The results of the coding highlighted the main themes under investigation in this study, Environment Setting, Auditory Elements (Environmental Audio and Auditory Feedback), Visual Elements (Visual Setting and Visual Feedback) and Controls. These themes also closely match the closed-ended questions asked as part of the quantitative analysis in Section 5.4.2.

Environment Setting

First, an investigation of participants' thoughts on the environment setting is explored. Each participant had different preferences for each environment, and the features under investigation impacted how they experienced the environment as a whole. We will highlight three topics that came up in the participants' feedback: "comfort", "focus", and "grounded".

Both environments provided different experiences of comfort when participants were asked, "what made you feel comfortable about the selected environment?". `P3` said that *"it felt more natural to walk along a forest path"*, and `P6` stated that the forest was *"probably more familiar"*. When talking about both environments, `P8` said that *"both were kind of familiar from various movies and video games"*. In the "space" environment, there is more of an expression of relaxation or a soothing experience from participants. `P5` stated when asked about the comfort in their preferred environment ("space"), *"not sure, felt very relaxing, could see myself putting the VR on to meditate in that space"*. While `P33` said that they *"find the vastness of space soothing"*. `P11` reflected on the both environments comfort saying, *"the locomotion in the forest environment was easy and comfortable, I also enjoyed the vegetation and bird sounds. the space environment was very peaceful"*. While the "space" environment itself provided a sense of peace, other aspects of the "forest" enabled comfort in different ways.

The participants used the phrase "focus" to describe their comfort in completing the pitch-colour/size tasks in the "space" environment. `P15` said there was *"much less "going on" so to speak. More minimal background was easier on eyes and enabled greater level of focus on task at hand"*. Similarly, `P38` said *"the space environment was more immersive and it was easier to focus on the task because the movements were easier and the environment was easier to navigate. the forest scene had more going on and looked less realistic"*, also the same participant mentions that *"not getting lost or losing a ball"* help them do the task in the "space" environment. While there was no mention of "focus" in the "forest", `P26`, whose preferred environment was the "forest path" said *"the experience was ok. I could easily be distracted by the surrounding environments and needed to close my eyes to concentrate at times"* when taking about the experience of doing the pitch-colour tasks in the "forest" environment.

An intriguing finding from the participants' feedback was their sense of grounding. This aspect was mentioned in relation to both environments, albeit in contrasting ways. In the "space" environment, three users noted the absence of a reference point or ground, with P16 stating, *"I am moving on the ground, so I would like to see a ground like Mars surface*, and P18 saying *"Some sort of reference point (plane or platform) in the space environment that would help with motion sickness."*. Also, P26 stated that they were uncomfortable with heights and that the glass floor in the tutorial gave them a bit of discomfort. With reference to the "forest", P26 mentioned that *"the forest path was more grounded and felt more believable. Therefore I engaged with it more"*. When referencing the comfort of the environment, P44 acknowledged that *"seeing the path ahead of me. It also felt brighter, less intimidating perhaps than space"*. While this is mentioned in terms of comfort, it can be that being able to see what is in front of you gives you a sense of being grounded.

Auditory Elements

This section breakdown the results from the template analysis into the two auditory elements explored in this study, the environmental audio that is used in the VE design, and the auditory feedback that is part of the cross-modal task between pitch and colour/size associations.

Environmental Audio In the "forest", the environmental audio was presented in 3D audio, while "space" had no environmental audio, but used a 3D audio for a location signal to find the next teleport pad. When asked what features the participants found most engaging in the VR experience, the 3D environmental soundscape was mentioned the most. P9 said *"For the forest path, the snap movement is quite jarring and made me feel a little nauseous but the soundscape is very relaxing and aids in the feeling of being there. In the space environment, the movement feels more natural but the soundscape takes away from the experience"*. When asked to compare the two environments in terms of immersion P41 said they *"felt the sounds in the forest were more immersive"*, while P13 mentioned that *"the soundscape of the forest seemed more immersive"*. This was a common thread among participants with P26 mentioning that *"the sound of environment and atmosphere in the Forest environment provided a stronger sense of presence of immersion*", and P40 said when referencing the "forest path" that *"the first environment seemed more familiar and the background sounds enhanced my experience"*. The "space" environment was not mentioned with regard to environmental sounds in itself but participants did comment on how peaceful it was, as mentioned in Section 5.4.4. In "space", the 3D sound used to locate the next teleport pad was highlighted as a very useful feature. P22 stated, *"in the Space environment, I liked the audio location helper to find the next place to move to."*, while P5 highlighted, when asked about what engaging features they liked, saying *"sound in space corresponded to the direction to turn"*.

Auditory Feedback Regarding the auditory stimuli from the pitch-colour/size tasks, the feedback centred around the lack of variation in the musical notes, instruments and timbre. While the audio stimuli were presented differently in terms of 3D vs 2D, this was not mentioned in the participants' feedback. P24 mentioned that *"it was good, though somewhat rudimentary. I found myself wishing for a wider*

palette of timbres and sounds". Several participants echoed similar statements when asked what improvements they suggest, with P17 saying they would like *"more instruments/sound fonts for the orb"*.

Visual Elements

This section presents participants feedback from the template analysis regarding the visual elements of this study. The two visual elements explored are the visual setting on the VE, and the visual feedback that is part of the cross-modal task between pitch and colour/size associations.

Visual Setting Analysis of the environments' visual setting provided a mixed response for both the "forest" and "space". Participants expressed lack of enthusiasm when talking about the "forest's" graphics. P15 sums up their experience with the following: the *"forest environment's graphics left a lot to be desired relative to space. Furthermore, the lag between balls in forest from walking as opposed to teleportation in space was slightly irritating"*. Similar sentiments were echoed by other users, for example, P9 stating *"There was some lag and virtual artefacts at the start of the forest. The forest was overall very pixelated when compared to the space landscape"*, and P34 mentioned *"more realistic quality in terms of animation might be helpful*. On a more positive tone, many participants did state that they found to "forest" to be immersive and fun, for example P4 stated that *"the forest path was more immersive than the space path, with more meaning given with the forest"*. In comparison, some participants mentioned how the "space" environment was visually pleasing, with P15 saying that the *"space VR location was particularly immersive and engaging. The quality and picture was enticing to look at"*.

Visual Feedback When commentating on the visual feedback of the pitch-colour tasks multiple references were made to the lack of colour choices when interacting with the musical orb. To highlight this, P41 suggested the use of *"finer color scale for orbs, increased range for orbs"*. P8 also felt the need for more colour options, suggesting that *"more colours to choose from and be able to walk on the grass. And maybe some deer"*, would improve their experience. With the secondary auditory-visual task involving pitch-size associations, not many participants expressed any insights on this. However, P44 stated that *"Sorting the colours and size of the orbs was a little boring to me, I could never really put a musical note to a size of orb in particular, without just moving along a scale from big to small, for example"*. Another participant (P29) found that size matched well with the auditory stimuli presented, explaining that *"Certain sizes felt more suited to certain colours and audio cues for me"*.

When asked what feature they found particularly engaging in each environment, participants mentioned the added visual feature in the "forest" where the fog changed colour in response to the participants' interaction with the orb. Many participants felt the colour-changing fog was helpful in the task, with P12 saying *"I felt that the forest environment allowed for more freedom to interact and the lighting effects were helpful to encourage task completion"*. To back up the previous statement, P17 said *"I found the more interactive environment (forest path) both easier*

to use and more fun. I really liked that when the colour of the orb was adjusted, the mist in the forest matched the colour. It was definitely more immersive".

While the orbs hue was the only colour feature that changed when the user interacted with it, participants could not find the right shade or brightness of that colour. This was highlighted when P9 said, *"I felt like I wanted more colour range in the orbs, my favourite colour is green but did not like the green in the environment"*, and P19 gained a new insight into their auditory-visual associations when stating, I *"discovered I don't really associate the colour red with sounds".*

Controls

When analysing feedback on the controls in the application, there were three features about which participants expressed their feelings: "movement", "turning", and "interaction controls". Many participants expressed a desire for the locomotion type from one environment to be available in their preferred environment, indicating that their favoured environment lacked the movement style they liked. P10 stated, *"Teleportation in the forest path might be more comfortable for me personally"*, while P12 expressed that they wanted *"more freedom to roam around in the space".*

There were other negative experiences with the two movement types within the different environments. In the "forest", participants did not like the fact that they were restricted to the pathway and could not enter the grass areas, and mentioned the movement interactions being rough and getting stuck when moving too close to the grass. P20 said, *"I could not walk on the grass in the forest, I would be stuck on a flower, which also altered the immersion".* At the same time, P18 mentioned, *"colliders. on forest pathway prevented movement at some turns so it meant adjusting position".* In the "space" environment, participants struggled to interact with the teleport pads when trying to point at the detection point on the teleport object. With the ray interactor pointers only appearing when the controller is engaged with the teleport pad, some participants had difficulty with this interaction. P33 said, *"sometimes it was hard to point at the teleport pads in the space environment"*, while P17 mentioned that *"the locomotion in the space environment was not great - it took a lot of flailing to get the pointer to aim at the teleport pad".*

Feedback showed that the turning controls used in each environment depended on personal preference. The snap-turning mechanic used in the "forest" was an unpopular choice with P9 saying *"for the forest path, the snap movement is quite jarring and made me feel a little nauseous".* However, some participants preferred it with P10 stating *"the jump-turning was a bit less disorienting than the continuous turning".* With regard to the continuous turning in "space", P11 mentioned that *"the locomotion in space (i.e.) continuous turning made me a bit dizzy, I preferred the snap turning in the forest environment".* Participants wanted the choice of turning style in each environment to feel more comfortable and combat dizziness. When talking about the "forest" P7 said, *"it caused me to get sick. This was down to the fact that I could not engage with the path in the way that I preferred".*

The final aspect of controls that was analysed was the interaction controls. Participants found that after time, they got comfortable with the controls, but for some, there was a steep learning curve. P22 with minimal VR experience said, it *"took a bit of time to learn the controls, but there was time to practice them before starting. I liked the interactions with size and sound and loved the coloured mist in the forest environment".* At the same time, P14 with no VR experience stated, *"I couldn't use*

the controllers correctly due to my own errors", and P39 expressed difficulty with the pitch-size interactions, saying it was *"hard enough to adjust the size. colours very easy to use"*. While P26 with over five years VR experience commented that the *"controls were very easy to use. The movement seemed fluid. The audio was of a high quality and immersive"*. Experience plays a significant role in how users engage with VR applications, showing the importance of providing participants with a simple sandbox/tutorial space before starting the main tasks.

One concern raised was the sensitivity of the controllers when making pitch-colour and pitch-size associations. P4 when referencing the "forest" environment said, *"the colour was not always easy to select (the environment goes very easily to pink and blue)"*. P20, offering suggestions on how to improve the experience, said *"reduce the shakiness of the hands+control. The controls pick up tiniest shaking of the hands which is then being translated into the game world. This reduces immersiveness and is annoying also reducing the precision when interacting with items in the world. I would suggest making this less sensitive and making items in the game (like buttons) bigger/ having more pixels so it's easier to aim."*. With a few participants skipping past questions in the tutorial and questionnaire scenes and issues with finer colour precision, controller sensitivity seems to be a small issue.

5.5 Discussion

Although no significant differences were found between the two environments across most quantitative metrics, the "forest" environment was associated with increased simulator sickness compared to the baseline (significantly), and "space". The findings are also reflected in participants' qualitative feedback, with common references to visual and movement-related issues. The following sections integrate both quantitative and qualitative findings to address the three research questions outlined in Section 5.1.

5.5.1 RQ1: How do the two VR environments impact user experience and performance in the pitch-colour association task?

Both environments provided good evaluation metrics but no significant differences in the SUS, IPQ and CLM scores (see Table 5.1). Compared to the baseline SSQ score, simulator sickness significantly increased in the "forest" and was higher than in "space", indicating that changes are needed to make the pitch-colour tasks more comfortable to take in VEs.

Positives and negatives can be taken from both environments. In terms of the optimal environment to engage with the pitch-colour/size association tasks, the "space" setting seems better suited for inducing focus on the task at hand. A study by Grimshaw-Aagaard and Walther-Hansen (2024) presented an argument against the *"more is more"* approach to VR that is more prevalent in VR applications at present, they state that *"where attention to the task at hand is required, presence in VR is desired, and stress is to be avoided"*.

Our results indicate that both environments provided a sense of comfort to participants. However, the "space" environment promoted greater focus for partici-

pants, but showed a higher cognitive load score (near significance) compared to the "forest". The "forest" was also viewed favourably but induced more stress, which was reflected in significantly higher SSQ scores compared to the baseline, and in the participants' feedback regarding issues with graphics and performance lag. These findings suggest improvements are needed in the VE design to lower cognitive load and reduce stimulator sickness, however in terms of environment setting for conducting pitch-colour association tasks, the a more minimalist VE would be suitable. With research highlighting links between reduced cognitive load and less visually complex VEs (Rettinger et al., 2024), issues with locomotion and controls (see Section 5.4.4) possibly contributed to the higher CLM score in the "space" environment.

With regard to completion times and variations in pitch-colour choices, no statistically significant differences were found between the two environments. However, as shown in Figure 5.4, participants tended to completed the task faster in the "space" environment, regardless of whether it was encountered first or second. Additionally, the average response time per note was lower for five out of the eight notes in the "space" environment, though the lowest (C4) and highest (C5) notes took longer to complete compared to the "forest" environment (see Figure 5.5). With previous research showing minimalist VEs being effective for learning and memory retention (Rettinger et al., 2024), this could have helped with the shorter times found in the "space" environment.

Several participants highlighted that the "forest" grounded them and "space" had no grounding and for some participants this was slightly uncomfortable. A study comparing a standard field-of-view (FOV) restrictor, where the user cannot see the ground, and a ground-visible FOV restrictor, showed the *"ground-visible FOV restriction offers benefits for user comfort, postural stability, and subjective sense of presence"* (Wu et al., 2021). This highlights an important feature and why it is important in VR design, while the IPQ scores were on average low for both environments, it was slightly lower for "space" (58%, compared to 59% for the "forest"). The term "grounded" does have a strong association with presence and what it represents, which is "being there", and by either positioning participants on a planet surface (as one participant suggested) or providing game mechanics to represent floating in "space", a higher sense of presence could be achieved.

With 69.2% of participants preferring the 3D audio design in the "forest", along with mostly positive feedback from participants when asked about "immersion" and "engaging features", it is clear that spatial audio helped to engage and immerse participants in the environment. Studies have shown that the use of spatial audio over stereo can result in increased performance in all of the UX (User Experience) sub-scales of presence, engagement, immersion, flow and emotion, making spatial audio a "necessary constituent of extended reality applications" (Kailas and Tiwari, 2021). Participants also gave positive feedback on the use of audio location in "space" when asked about the features they found engaging. This feature was not used in the "forest" as the participant could see their next target right in front of them, while in "space" new targets appeared in different locations after each interaction. While this feature worked well studies have shown that auditory-visual cues are more effective than audio alone at helping the user locate objects (Broderick et al., 2018). This audio location feature was also set up to help combat cognitive overload, as giving participants cues or signalling has been shown to increase learning performance (Mutlu-Bayraktar et al., 2019). A single preferred environment between

the two investigated was not found but effective features from each environment will provide valuable guidance in designing future VR environments.

5.5.2 RQ2: How do changes between the two environments affect participants' experiences?

The different implementations of locomotion, audio and visual elements had a big impact on the participants' experiences, with a few of these explored in Section 5.5.1. While there was no significant difference between the two environments from the standardised assessment scales of the `SUS`, `IPQ`, and `CLM`, a lot of interesting insights were gained from these tests and also from the template analysis.

The analysis of the `CLM` showed a marginal significant difference in cognitive load between the two environments, but with a moderate level of cognitive load observed. With "space" scoring 58.55% and the "forest" scoring a 50.33%, participants found the "space" environment slightly more cognitively taxing. This suggests that visually rich VEs, such as the "forest", do not necessarily increase cognitive load, as factors like environment settings, audio presentation, movement, and navigation can help mitigate its effects. Two key factors may contribute to reducing cognitive load. First, the type of environment plays a role, as research has shown that virtual forests can support cognitive recovery (Lee et al., 2022). Second, effective multi- and cross-modal design has been found to help reduce cognitive load by distributing cognitive processing across multiple sensory channels (Bu et al., 2025). Given the importance of managing cognitive load to enhance users' ability to form cross-modal associations (Morrison et al., 2015), the findings of this study provide insights for designing future VR experiences that minimise cognitive strain and optimise learning outcomes

The movement and turning in both environments was one of the major pain points brought up by participants, with locomotion in the "space" environment slightly preferred with 53.8%. Issues were found with both movement (free and teleportation) and turning (snap and continuous), as some participants experienced motion sickness. However, no single locomotion type solely accounted for it. While a lot of different movement types are available in VR experiences, such as walk-in-place, leaning, arm swinging, tapping-in-place, joystick (used in study) and teleportation (used in study) (Cherni et al., 2020), the last two are popular artificial locomotion methods that work well in limited physical spaces (Boletsis and Cedergren, 2019). One of the drawbacks to using joysticks in immersive VR applications is that it can cause some disorientation (Wilson et al., 2016), with teleportation having been found to induce minimal motion sickness compared to other types such as free movement (Cherni et al., 2020). With the amount of variability in the preferred movement and turning styles across the study, giving the participant choice between the two movement and turning types or implementing both into the design at the same time would allow participants to engage with the environment in a way most comfortable for them, therefore providing a higher sense of presence.

While there were not too many insights on the difference in audio feedback in the pitch-colour association tasks, with the audio presented in 3D in the "forest" and 2D in "space", the visual feedback provided some interesting results. Though subtle, the added colour interaction with the fog in the "forest" was highlighted as a feature participants found engaging, providing increased immersion and also helping some people focus on the task. This feature was included to give participants more of

an experience of what a sound-colour synaesthete could possibly experience, while not an exact representation, as this experience is individualised to each person. Curwen (2018) states that comparisons between non-synaesthetes and synaesthetes in cross-modal correspondences (e.g., colour, music, emotion, pitch-height, pitch-size) suggest similar mental processes. And that synaesthetes may not be inherently predisposed to synaesthesia but might develop unique abstract concept processing through their synaesthetic "extra qualia" (Curwen, 2018; Wager, 1999). If this is the case, can mimicking this synaesthetic "extra qualia" in immersive VR experiences help support a stronger understanding of musical stimuli and aid in learning simple music concepts? Participants mentioned how this visual feature provided more immersion and helped with the task when compared to the other environment that did not provide this extra visual feature. Also, several participants mentioned that there should have been more colour choices, which shows that maybe some people have a stronger association between pitch and colour that is more synaesthesia-like, as mentioned in Section 5.2.1.

5.5.3 RQ3: What improvements would make the system more effective for the pitch-colour association task?

The results have given informative insights into how this VR experience can be improved to provide a more complete, engaging, immersive environment better suited for pitch-colour association tasks. The previous two questions already indicated some improvements that would make this system more effective, in terms of locomotion options, audio design, and improved visual settings. Next, we will discuss a few more important suggestions, including colour choices and control interactions.

Seven participants expressed issues with the limited range of colours, suggesting that some people may be more attuned to associations between their senses and were frustrated when they couldn't find their colour choice. Also, some participants (see 5.4.4) discovered that certain colours did not associate with pitch for them. This shows a new learning for the participant, and shows that while more generalised cross-modal associations provide a good bases for our understanding of pitch-colour associations, it can be a more personal association for some people, even those who are not deemed to have synaesthesia. As mentioned in Section 5.2.1, this reinforces previous research that suggests looking at synaesthesia (in our case, pitch-colour synaesthesia) as not a binary ability but a spectrum (Campen, 2010).

From analysing the frequency of pitch-colour choices in Figure 5.6, a slight trend emerged: lower notes were more frequently associated with darker colours, while higher notes showed a similar, though less pronounced, tendency toward brighter colours. This pattern aligns with previous research on pitch-colour cross-modal correspondences (Spence, 2011; Ward et al., 2006). Since VR enables real-time adaptation of visual feedback, it presents an opportunity to refine these pitch-colour associations dynamically. Expanding on this, the final application design could incorporate a more responsive colour-changing system, allowing for both simple and advanced visual feedback mechanisms to accommodate variations in user perception. To maintain clarity and avoid overwhelming participants, the colour options in the task were based solely on hue. Previous research suggests that single musical notes tend to elicit basic emotional and colour associations rather than more complex interactions (O'Toole et al., 2019), reinforcing the decision to limit the complexity

of pitch-colour mappings in this study.

In the secondary task, participants were asked to make pitch-size associations alongside pitch-colour associations. However, only 83% of objects were successfully triggered for size manipulation, suggesting that this additional task may have been cognitively demanding for some participants. As noted in Section 5.4.4, some participants found the object sizing mechanic difficult to use, whereas the colour-changing mechanic was perceived as more intuitive. The limited selection of "smallest" and "smaller" size categories, along with the incomplete adoption of size manipulation, suggests that issues with the mechanic may have contributed to difficulties in fully engaging with the pitch-size association task. Also, while incorporating multiple cross-modal features between auditory and visual modalities can enhance our understanding of perception and how immersive VR can harness these relationships, presenting multiple cross-modal tasks simultaneously may have exceeded participants' cognitive processing capacity. Despite this, Figure 5.7 highlights interesting insights, particularly the higher frequency of lower pitch-larger size associations. This finding supports the presence of pitch-size cross-modal associations in VR environments and aligns with previous research (Evans and Treisman, 2010; Gallace and Spence, 2006; Spence, 2011).

Another suggested improvement to the VR experience concerned the controller interactions and sensitivity. Several participants mentioned the sensitivity of the controllers in a few instances, but particularly when trying to make the pitch-colour/size associations while interacting with the musical orb. There may be a slight overlap between the controller interaction to change the colour and the issue of not enough colour choices mentioned in the previously. A couple of references to the colour interactions were slightly ambiguous, like `P4's` quote mentioned in Section 5.4.4. This quote could be in reference to the control sensitivity or finding certain colours and a need for finer choices. The precision of the `Oculus` controllers has been tested and deemed suitable for medical use, with high precision found (Pereira et al., 2023). Adjusting the parameters for the colour change interaction might help fix this issue, along with using the "x" and "z" axis for more detailed colour changes.

Based on the results of this study, no clear preferences from the two environments presented here can be found. However, invaluable insights from the thirty-eight participants was garnered allowing us to get a sense of what is important in creating a comfortable immersive VE suitable for focusing on an pitch-colour association tasks that allows the user to be engaged and present within the VR space. The suggested improvements stated above, look to improve certain features from each environment, creating more choice for the user in how the engage with the task presented to them. Creating a single environment that is suited to task at hand, but has enough personalisation to make the users comfortable and engaged would be an interesting design challenge. It seems that with the variability of preferences on different features, a lot more has to be taken into account to provide an optimal environment for each person. Future studies could investigate how learning styles and personality can help in adapting the environment and interaction styles to better suit different individuals.

The insights from this study highlight how virtual environments (VEs) can facilitate immersive and focused learning experiences by leveraging auditory-visual cross-modal associations. Outside of music, this approach has broader applications

across various educational domains, such as language learning and special education. Extended Reality (XR) technologies, encompassing Virtual Reality (VR) and Augmented Reality (AR), have been shown to enhance language acquisition (Chen et al., 2022). By integrating cross-modal associations—such as linking visual cues with auditory language input—learners can engage in contextualised, interactive experiences that improve comprehension and retention. Cross-modal associations in XR could enhance mental health interventions by dynamically adapting virtual environments to users' affective states. Bermudez i Badia et al. (2019) demonstrate how multi-modal representations of emotions in VR can improve emotional self-awareness and regulation. Personalised cross-modal mappings could refine this approach, aligning sensory stimuli with individual emotional responses to support anxiety management, mood regulation, and stress reduction in immersive therapy.

By harnessing cross-modal associations in XR design, educators and developers can create more effective and inclusive learning tools that cater to diverse educational needs.

5.6 Limitations & Future Work

There are a few limitations to highlight from this study. The standardised assessment scales of `SUS`, `IPQ`, `CLM`, and `SSQ` require time to complete after each environment, potentially affecting the immediacy and accuracy of the scores (Schwind et al., 2019). Although short-form scales were used to mitigate this, the delay remains a potential limitation.

The template analysis highlighted potential technical limitations of the `Oculus Quest 2`, particularly in how it rendered the visually rich natural "forest" environment. Some participants noted that the graphical quality of the "forest" setting was inferior to the "space" environment and suggested improvements in visual fidelity. Specific concerns included objects unexpectedly appearing in the scene, such as stones materialising in ways that disrupted realism. These graphical limitations stemmed from the challenge of rendering a realistic forest on a stand-alone headset, leading to a more stylised aesthetic and occasional graphical artifacts. While running the application on a VR-capable computer with high-powered graphics could enhance visual quality, this approach would restrict participant mobility and limit where the study could be conducted.

Improving performance using realistic graphics and increasing immersion and presence will be important to engage and focus the user in the pitch-colour/size tasks. A good balance between the *less is more* approach (Grimshaw-Aagaard and Walther-Hansen, 2024), and need for high quality graphics for full immersion (Silva Bastos et al., 2018) should be considered. Although this study explored only two virtual environments, the concept of learning environments in VR warrants further investigation. The "forest" and "space" environments were selected for their contrasting characteristics, each offering unique elements to create a comfortable learning atmosphere. Future research should, however, examine additional environments to broaden the understanding of their impact on learning.

Building on the insights gained from this study, future research could take a more quantitative approach by isolating single factors and assessing their specific impact on pitch-colour cross-modal associations in VR settings. While this study intentionally used contrasting virtual environments to explore holistic design effects,

multiple elements—such as environment aesthetics, spatial audio, visual effects, and locomotion—were varied simultaneously. This multifactorial manipulation limits the ability to attribute effects to any one feature. Future studies would therefore benefit from systematically varying one design element at a time, allowing for clearer causal interpretations.

Such an approach would further refine our understanding of how to optimise the design of virtual applications that incorporate pitch-colour cross-modal interactions. Research could, for instance, isolate specific elements such as locomotion or audio spatialisation, and evaluate their effects using objective behavioural or physiological measures.

Additionally, exploring other individual auditory and visual features—such as pitch-size, loudness-colour, or timbre-colour—would offer a deeper understanding of auditory-visual cross-modal associations in VR and how they can be implemented into immersive learning environments. Given the complexity of music as a multi-modal system, advancing beyond pitch-colour relationships to examine the interplay of multiple auditory-visual features may lead to the development of personalised cross-modal learning systems in VR. Such an approach could enhance usability and engagement, ultimately improving the effectiveness of VR-based music education and training applications.

This study has provided insights into various VR features that can inform a cross-modal design process for creating immersive and engaging virtual learning environments. With the pitchy-colour/size tasks participants' undertook in this study, a virtual musical learning environment could be explored, while also showing benefit in other educational fields. It is important to understand what is needed for the task at hand and to focus on implementing these features to a high standard to achieve the goals of the research, which is an effective cross-modal VR musical learning experience.

5.7 Conclusion

The aim of this study was to compare two virtual environments that present different features related to, environment setting, audio presentation, locomotion type and auditory-visual feedback, to see how participants engaged with the environment while completing an pitch-colour and pitch-size association task.

Thirty-eight participants followed an A/B design providing feedback on both environments through the standardised assessment scales of `SUS`, `IPQ`, `CLM`, and `SSQ`, and also filling in a post-study questionnaire. The quantitative analysis provided few insights on variances in colour choices, the completion time, and the preferred environment in terms of usability, presence, and cognitive load, however a significant increase in simulator sickness was found across all the sub-scales in the `SSQ` for the "forest".

The template analysis provided rich insights which highlighted the pains and gains for different features related to the application's environment setting, auditory elements, visual elements and controls. Participants' opinions on a preferred VR environment varied. However, in terms of environment setting, the more minimalist "space" seemed to focus participants on the pitch-colour tasks. Other important insights were found on locomotion, with motion sickness symptoms an effect from both movement types. The 3D audio presentation in the "forest" was seen in both

the quantitative and qualitative analysis to be preferred and provide a greater scene of immersion, while the extra synaesthetic visual feature in the "forest" helped with task focus and immersion in the environment.

Another interesting insight was that many participants felt the colour they associated with the musical pitch stimuli was not in the range of colours provided. This leads to a suggestion that some participants have stronger associations between pitch and colour, with some happy to use more general colour options (hue) and other needing finer choices (adding saturation and value). The insights gained from this study offer a foundation for enhancing cross-modal design in VR applications, particularly in the context of pitch-colour association tasks. These findings not only inform the development of immersive experiences but also pave the way for future research into leveraging VR for intuitive and engaging learning platforms, where cross-modal interactions can be tailored to diverse educational and creative applications.

5.8 Chapter Reflection

Continuing the exploration of this thesis's central question—*"How can the cross-modal associations of pitch and colour be enhanced through virtual reality?"*—this chapter focused on the sub-question presented in Section 1.3: *"What role do immersive features play in supporting pitch-colour cross-modal associations in VR learning environments?"*. Building on the findings from Chapter 4, which showed no performance difference between desktop and non-immersive VR settings, this study shifted attention to the role of environmental design, spatial audio, interaction, and user control in enhancing user experience within a pitch-colour cross-modal task.

Using a mixed-methods approach, the study employed an A/B test between two different environment and feature settings to explore how users engaged with the pitch-colour association task. Valuable insights were drawn from both quantitative and qualitative data. Quantitatively, the more visually complex "forest" environment resulted in significantly higher simulator sickness scores compared to the baseline, with contributing factors including visual busyness and locomotion style. Regarding pitch-colour associations, the variance in user responses was similar between environments, though a trend emerged aligning low pitches with darker colours and high pitches with brighter colours.

The qualitative template analysis provided a richer understanding of the user experience. Participants generally preferred the "space" environment, citing its relaxed atmosphere and task-focused design. Insights also emerged around locomotion and visual feedback: while no single locomotion style was universally preferred, participants appreciated the synaesthetic-like visuals in the "forest" environment, even if the overall experience was more overwhelming.

The findings highlight several considerations for designing with cross-modality in mind. Selecting an appropriate environment is essential—not only to create immersion but also to maintain user focus, comfort, and retention. Moreover, the perceptual nature of cross-modal associations means that individual variability must be acknowledged. While the observed trends reflect findings from prior research on pitch-colour associations, some participants expressed frustration that their preferred colour choices were not available. As discussed in Section 2.2.3, the line between cross-modal associations and synaesthesia remains debated, with some re-

ferring to "weak synaesthesia" in this context. This points to the need for adaptable mapping systems that can accommodate individual differences in sound-visual associations.

Although the results from the studies in Chapter 4 and this chapter are not directly comparable in terms of pitch-colour associations, the investigation into immersive features has informed design considerations for cross-modal VR environments. Furthermore, as highlighted in Section 2.4.1, and reinforced by this study, high technical immersion (e.g., realistic graphics or detailed audio) is not always necessary. The preference for the more minimalist "space" environment suggests that a simpler design can often support user engagement more effectively than one that attempts to replicate the physical world in detail.

In addressing the central question—*"How can the cross-modal associations of pitch and colour be enhanced through virtual reality?"*—this study has provided valuable insights into how design elements such as environment, locomotion, controls, and sensory presentation influence user engagement and experience. The three pillars of immersive VR outlined in Section 2.4—immersion, presence, and interaction—play a significant role in creating meaningful experiences that support cross-modal learning. This interplay must be carefully balanced, considering both positive and negative impacts on the user. As discussed in Section 2.5.1, the link between cross-modal associations and multisensory integration makes VR an ideal platform for delivering integrated sensory experiences that can strengthen perceptual links across modalities.

This study represents an important step towards understanding how to design VEs that support the integration of auditory and visual senses, by identifying how to structure VR spaces that allow users to focus, engage, and learn through cross-modal interaction.

The insights gained here directly informed the final study in this thesis, presented in Chapter 6. That study builds on the feedback and findings of this chapter to present a unified VR learning environment designed for music education. A single "space" environment with improved locomotion options is used to explore how participants can learn basic music theory using pitch-colour associations. Responding to earlier findings around both general and personal colour mappings, the next study introduces a comparison between a predefined mapping—based on prior research—and a personalised mapping developed by the users themselves.

Chapter 6 aims to generate further insights into the application of VEs for supporting cross-modal learning. It also grounds the research in a practical use case: teaching basic music theory. This allows for the investigation of how pitch-colour cross-modal associations might be applied in real-world educational contexts.

Chapter 6

Exploring Cross-Modal Pitch-Colour Associations for Music Learning in Immersive Virtual Reality

This chapter is based on a study that has been submitted to the *Computers & Education: X Reality Journal*, under the title: P. O'Toole, M. Mancini, I. Pitt, and Laura Maye. Exploring Cross-Modal Pitch-Colour Associations for Music Learning in Immersive Virtual Reality.

The following study explores the sub-question presented in Section 1.3: *How Do Personalised & Pre-Defined Pitch-Colour Cross-Modal Mappings Compare in Supporting Learning Outcomes in VR-Based Music Exercises?*

Abstract: *Cross-modal mappings between pitch and colour offer promising avenues for enhancing perceptual understanding in music-related learning. Virtual Reality (VR) provides a unique platform for exploring such associations and their potential in learning musical theory. This study explores the use of pitch-colour associations within an immersive VR music learning environment. Specifically, it examines how pre-defined pitch-colour mappings (those where colours are already mapped to a specific note) and personalised mappings (those where participants choose which colours are mapped to a specific note) can support music learning tasks in VR. Thirty participants were divided into two groups—one using pre-defined mappings and the other creating personalised ones. Each group completed eight music lessons based on the C Major Scale using a VR headset. The between-subjects design examined whether personalised mappings support learning-related performance and user engagement, compared to the pre-defined group, using both quantitative measures (error rates and task completion time) and qualitative template analysis. No statistically significant performance differences were found between groups, although the small sample size may have limited the power of the quantitative analysis. In contrast, the qualitative data provided valuable insights into participants' experiences, highlighting how both mapping strategies shaped engagement and understanding with the music lessons. Notably, significant improvements in error scores and*

completion times were observed across the eight music lessons for both groups. Additionally, VR design elements—including locomotion, interaction, and multimodal cues—were found to influence immersion and usability. These findings demonstrate the potential of VR applications that use a pitch-colour cross-modal mapping design to support music-related tasks such as interval recognition, pitch differentiation, and ear training. They also inform future design considerations for cross-modal interaction and its potential in immersive learning environments.

6.1 Introduction

In today's rapidly evolving technology-driven society, Virtual Reality (VR) is increasingly explored as a viable learning tool (Lorenzo et al., 2022; Marto et al., 2022; Trujano et al., 2018). Notably, VR has the ability to present multiple sensory cues simultaneously, facilitating multimodal learning approaches that align with the brain's natural integration of information, resulting in more efficient and robust skill acquisition compared to uni-sensory methods (Shams and Seitz, 2008). VR also enables precise environmental control, allowing researchers to present either minimalistic, distraction-free settings or rich, affective environments tailored to the task—conditions that are difficult to replicate in the physical world (O'Toole et al., 2025; Riva et al., 2007).

VR's full potential remains largely untapped, particularly in terms of creating engaging 3D environments, incorporating cross-modal sensory integration, and providing immersive interactivity for designing music learning applications. This highlights a significant gap in both research and development, suggesting that VR could offer transformative opportunities to advance music theory education through novel, interactive experiences that traditional digital tools cannot provide.

Given that VR is inherently a multimodal technology, its potential for cross-modal interactions remains relatively unexplored in VR application design. Multimodal interactions integrate distinct sensory modalities (e.g., visuals, sounds, and haptics) to create a coherent, immersive experience, such as seeing a virtual instrument and simultaneously hearing its notes (Jacko and Sears, 2003). In contrast, cross-modal interactions occur when one modality automatically evokes or reinforces another, for instance, when a specific pitch triggers a corresponding shift in colour, thereby deepening pitch-colour associations (Spence, 2011). A study by Malpica et al. (2020) found that cross-modal effects are present in VR environments using head-mounted displays (HMD), in a similar way to conventional displays. Although some research into cross-modal design in VR has been conducted—such as studies on Reduced Realities (RR) (Grimshaw-Aagaard and Walther-Hansen, 2024), classical musical narratives in XR (Rudenko et al., 2022), and delivering selective high-fidelity environments (Chalmers et al., 2009)—its application to music education remains relatively unexplored.

This study seeks to understand whether pitch-colour cross-modal associations, presented in an immersive VR environment, can support learning scales and intervals based around the C Major Scale. In terms of learning support, this study focuses on error rates, task completion time, and qualitative feedback on users' experiences, with any perceived enhancement of learning-related outcomes interpreted in an exploratory manner.

Eight different lessons were presented to participants, including an ascending scale, a descending scale, three individual intervals (a third, a fifth, and an octave), and three different interval-based scales (see Figure 6.4). These lessons were chosen because the C Major Scale does not have any accidentals (sharp or flat notes) and are commonly used for completely beginners to music theory [1].

To explore the research questions, thirty participants were split into two groups: fifteen in the control group (CtrlGr), who used a pre-defined pitch-colour mapping where specific colours were already assigned to each note, and fifteen in the experimental group (ExpGr), who created their own personalised pitch-colour mappings by selecting which colours they felt matched each note before taking part in the music test. Previous research on pitch-colour cross-modal associations informed the pre-defined mappings used by the CtrlGr in this study. This research suggests that higher-pitched notes are commonly linked with brighter colours and lower-pitched notes with darker ones (Spence, 2011; Calvert et al., 2004; Martino and Marks, 1999; Marks, 1987). The personalised pitch-colour mappings that ExpGr participants create are inspired by sound-colour synaesthesia, where individuals experience consistent, involuntary colour associations with specific sounds or pitches (Ward, 2013; Curwen, 2018).

In this study, personalised mappings refer to each participant's own chosen colour for each pitch, based on their immediate cross-modal impressions. Exploring both mapping strategies allows the study to compare the benefits of structured, generalisable associations against the potential experiential and perceptual advantages of personalised, intuitive links. This is particularly relevant in VR learning contexts, where immersive and adaptive design can synchronously present auditory and visual stimuli, engaging multisensory integration mechanisms that may reinforce cross-modal associations and support perceptual learning and memory (Shams and Seitz, 2008; Spence, 2011).

This study hypothesised that the Experimental Group would achieve lower error scores and faster completion times than the Control Group during the music lesson tasks, indicating more effective performance. To investigate this, performance differences between pre-defined cross-modal mappings (CtrlGr) and personalised mappings (ExpGr) were explored through the following research questions:

- RQ1: In an immersive VR cross-modal music test, how does participants' performance differ between the control and experimental groups?
- RQ2: Does participant *"performance"*—defined by error scores, completion times, and reported learning experiences—improve over the eight lessons for both groups?
- RQ3: What kinds of pitch-colour associations did participants of the experimental group make?
- RQ4: How can immersive virtual environments for musical education improve going forward?

In addition to examining how each group performed in the music tests, this study explored which aspects of the immersive VR experience participants found

[1] Do you want to major in music?,`https://iastate.pressbooks.pub/majoringinmusic/chapter/music-theory/`, accessed on January 29, 2026

most beneficial when engaging with the learning task. These aspects included the environment, locomotion, and interactions. Both qualitative and quantitative data were collected, with template analysis employed on the qualitative data. Through these combined methodologies, the research aims to advance our understanding of how best to design musical learning scenarios in VR by harnessing cross-modal associations. Given the need for music education to evolve beyond traditional models, this study demonstrates how VR experiences can support engaging, personalised, and interactive forms of learning, using a platform that appeals to younger generations.

6.2 Background

6.2.1 Pitch-Colour Cross-Modal Associations

Cross-modal associations (explained in Section 6.1) between the auditory and visual senses (also known as sound-colour cross-modal associations) can occur between several different musical features, such as music key, timbre, melody, and pitch, with visual features such as colour, shape, size (Spence, 2011; Martino and Marks, 1999; Calvert et al., 2004; Adeli et al., 2014; Palmer et al., 2013).

Because pitch (an auditory attribute) and colour (a visual attribute) are central to this study, the following review of prior research will emphasise cross-modal associations between these two features. It has been observed that pitch associates with brightness, lightness in the visual sense (Spence, 2011; Calvert et al., 2004; Martino and Marks, 1999; Marks, 1987). These associations have been found to associate higher-pitch sounds with brighter, lighter colours (Spence, 2011). Several studies have also shown that emotion can influence colour choices for musical features such as melody and pitch (Palmer et al., 2013; O'Toole, 2021), suggesting that affective factors may shape how these cross-modal connections form.

Sound-colour synaesthesia shares commonalities with sound-colour cross-modal associations, with differences in the automaticity, consistency and specificity of the feature choice (Ward, 2013; Curwen, 2018). Sound-colour cross-modal associations have been extensively studied in experimental psychology, both in synaesthetes (Cytowic, 2003; Ward et al., 2006; Farina et al., 2017; Spector and Maurer, 2009; Sun et al., 2018), and the general population (Sun et al., 2018; Spence, 2011; Martino and Marks, 2000). Synaesthesia typically arises when one sense is stimulated (an inducer) and it triggers a reaction in an unstimulated second sense (a concurrent) (Cytowic, 1989; Curwen, 2018). It is prevalent in about 4% of the general population, with sound-colour synaesthesia (also known as chromaesthesia) being less prevalent at 0.2% (Curwen, 2018; Farina et al., 2017).

Research into synaesthesia, using tools like the `Synaesthesia Battery` (Eagleman et al., 2007) and the `Test of Genuineness (TOG)` (Asher et al., 2006), suggests it may not be a strictly binary condition. Campen (2010) observed variability in consistency test scores, with some non-synaesthetes showing strong sound-colour associations, proposing that synaesthesia exists on a spectrum rather than as a distinct classification.

Taken together, this literature shows that while pitch-colour pairings are common across people, they also vary in consistency and strength depending on individual factors. This highlights an opportunity to explore how immersive VR environments

could help reinforce or expand these connections by enabling users to engage with auditory and visual cues through spatial interaction, real-time feedback, and personalised sensory mappings. The following sections, therefore, examine how VR can use cross-modal design in a music learning context and what specific features may support pitch-colour associations.

6.2.2 VR as a Learning Tool: Designing Immersive and Cross-Modal Environments

Extended Reality (XR)—encompassing Virtual Reality (VR) and other immersive technologies—has gained increasing attention across healthcare, gaming, music, and education (Malpica et al., 2020). Research suggests that VR's capacity for experiential learning and high interactivity can enhance motivation and engagement in educational contexts (Mantovani, 2001). At the same time, carefully designed VR environments can elicit affective responses that align with particular learning goals (Riva et al., 2007), reinforcing VR's role as an affective medium.

Evidence of VR's Efficacy in Learning

Several studies confirm VR's advantages in education, highlighting the importance of environment design, locomotion options, and cross-modal strategies. Ou (2023) reported that multimodal music instruction led to significantly higher scores in rhythm, musicality, improvisation, and memory compared to traditional methods. In a Kindergarten to Sixth Grade (K-6) setting, a meta-analysis demonstrated that immersive VR provides greater learning benefits than semi- or non-immersive systems (Villena-Taranilla et al., 2022).

As highlighted in Section 2.4.1, immersion does not always require high-fidelity experiences; instead, a more nuanced approach to immersive design can be equally effective. The *"less-is-more"* approach proposed by Grimshaw-Aagaard and Walther-Hansen (2024) illustrates how reduced realities can minimise sensory overload, reduce stress, and open new possibilities for immersive learning scenarios. This aligns with studies by Young et al. (2025) and Makransky et al. (2019), which suggest that although high-immersion environments can increase presence and motivation, they may also introduce greater cognitive load, potentially hindering learning outcomes if not carefully managed.

In summary, effective VR design for learning contexts require a careful balance, one that considers not only the potential of immersive features but also their cognitive demands. Research-driven approaches that integrate suitable environments, intuitive interaction, and congruent cross-modal feedback can facilitate learning experiences, especially when designed to minimise overload. Ongoing investigation into how immersion, user interaction, affective design, and sensory congruency interact will further refine immersive tools that are both engaging and pedagogically effective.

Virtual Environment Design and Learning

Environment choice can strongly influence user experience and certain cognitive and learning-relevant processes. Natural scenes (e.g., forests) reduce stress and negative emotions, whereas urban or highly stimulating environments may cause fatigue and

lower self-esteem (Wang et al., 2019; Yu et al., 2018; Lee et al., 2022; Mattila et al., 2020). Minimalist settings limit distractions and aid focus, memory, and spatial recall (Gkoumas and Shimi, 2022). In a comparison of a *"forest"* and a simplified *"space"* environment, O'Toole et al. (2025) observed that users preferred the minimalist design for complex auditory-visual tasks, indicating that simple design can foster focus.

Affective design further shapes learning-related experiences in VR. By incorporating specific emotional tones, designers can influence presence and engagement, which in turn affect learning (Riva et al., 2007). Likewise, combining VR with affective computing—where virtual environments adapt to users' emotional states—enables more personalised and dynamic educational experiences (Ontiveros-Hernández et al., 2013).

Locomotion is a critical factor in VR, directly affecting immersion and comfort. Boletsis and Cedergren (2019) compare different locomotion methods, noting that walking-in-place offers high immersion but can be uncomfortable, controller-based movement is user-friendly but less realistic, and teleportation reduces motion sickness while disrupting visual continuity. Cross-modal design further influences both comfort and potential learning experiences, particularly through vection—the illusion of self-motion. Kim et al. (2023) found that aligning auditory and visual cues helps delay discomfort, whereas mismatched cues can intensify motion sickness.

6.2.3 Musical Education

Music theory, at a basic level, can be viewed as the *"grammar"* of music, helping understand concepts like key signatures, scales, intervals, and chords. The study of more advanced theory topics (such as harmony, counterpoint and form) leads to a deeper understanding of the inner workings of music [2]. Music notation in the Western tradition, although invaluable for notating complex compositions, can feel daunting to newcomers. Pedagogical methods such as `Orff`, `Kodály`, and `Dalcroze` leverage multimodal approaches that combine auditory, kinaesthetic, and visual elements, such as singing, movement, and improvisation, to nurture intuitive engagement with pitch and rhythm [3]. Central to these approaches is `solfège` (Do-Re-Mi), which aids in interval recognition by mapping syllables onto scale degrees. Additionally, the circle of fifths visually represents harmonic relationships, illustrating how musical keys in the chromatic scale connect via intervals of a fifth.

Historically, several theorists have extended basic learning methods by coupling sound and colour. In 1704, Isaac Newton proposed a colour-music wheel linking the seven colours of the visible spectrum to the diatonic scale, conceptualising rising musical pitch as paralleling increasing optical frequency (Greated, 2011). Later, Alexander Scriabin, believed to be a synaesthete, created a colour-music system for *Prometheus: The Poem of Fire (1910)*, anchoring each note in the circle of fifths to a specific hue (see Figure 6.1). While both Newton and Scriabin assigned red to *"C"*, there has never been universal agreement on these mappings, reflecting the subjective nature of sound-colour correspondences. In recent work, O'Toole

[2]Serenade Magazine,`https://serenademagazine.com/music-theory-pedagogy-how-its-taught-and-why-it-matters/`, accessed on January 10, 2025

[3]Chordblossom,`https://www.chordblossom.com/writing/music-teaching-methods`, accessed on January 10, 2025

et al. (2019) found that participants often associated darker colours, like blue, with lower pitches, and brighter colours, such as cyan or yellow, with higher pitches, aligning with research linking higher tones to lighter or brighter colours (Spence, 2011; Calvert et al., 2004; Martino and Marks, 1999; Marks, 1987).

Building on these ideas, a sound-colour system can potentially serve as a novel way reinforce understanding of intervals and scale practice, especially when combined with immersive learning experiences. By actively engaging both auditory and visual senses, cross-modal design may foster deeper conceptual understanding, increase motivation, and make music theory more accessible by allowing learners to approach it through multiple sensory perspectives. Giving participants the ability to personalise their pitch-colour associations may strengthen this effect, as self-chosen mappings could feel more intuitive and meaningful than pre-defined ones. This personal connection might help with memory and focus during the task, potentially supporting more effective task-based learning than using a fixed mapping.

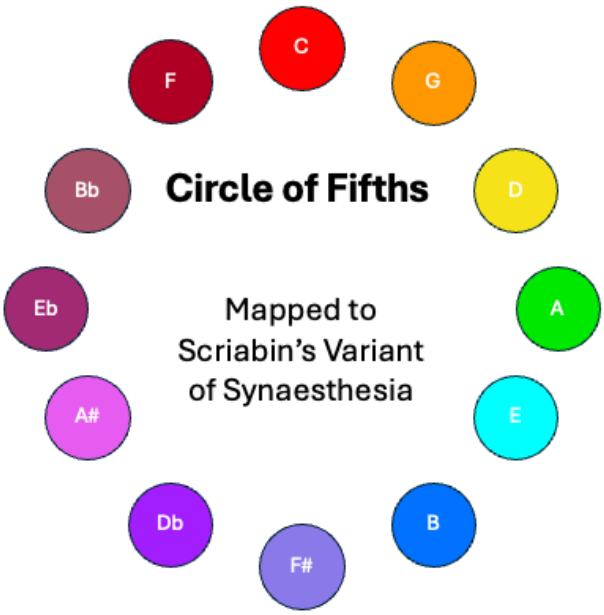


Figure 6.1: In music theory, the circle of fifths organises the 12 notes of the Western scale in ascending intervals of a fifth, showing their key signatures and relationships. Composer Alexander Scriabin famously extended this idea by associating each key with a specific colour; his sound-colour associations are shown here. The image is based on an image in Rucsanda (2017).

6.2.4 Current VR Music Applications

While VR has gained popularity across multiple domains, its application in music education and structured music learning remains limited. Commercial VR games such as `Beat Saber`[4], `Rock Band VR`[5], and `Pistol Whip`[6] have been highly successful, but they focus predominantly on rhythmic gameplay rather than structured musical training.

[4]Beat Saber, `https://beatsaber.com/`, accessed 10 August 2024.
[5]Rock Band VR, `https://www.rockbandvr.com/`, accessed 15 January 2025.
[6]Pistol Whip, `https://www.cloudheadgames.com/pistol-whip`, accessed 15 January 2025.

Some recent VR applications, including `Virtuoso VR`[7], `PatchWorld`[8], and `VRTuos`[9], support creative musical experiences. However, these platforms are primarily oriented towards music creation and improvisation rather than systematic theory or ear-training instruction. Educational tools such as `PianoVision`[10] and `Perfect Pitch`[11] introduce limited structured content like basic piano or pitch recognition exercises, but rarely leverage VR's immersive and cross-modal potential for deeper musical understanding. Thus, this study aims to address a notable gap in utilising VR for structured music theory and perceptual training.

Music Learning in VR

Research on integrating music learning into VR environments has explored harmony and chord progressions, the use of spatial audio, visual cues, and musical instrument learning. One example is a study by Timoney et al. (2018, (p. 1)) that proposed the *"Singing Blocks"* application, which uses the *"innately physical nature of VR"* to make learning harmony and chord progressions more immersive—a level of engagement they argue is *"impossible with other approaches"*. Although no formal study followed this proposal, it underscores VR's potential as an effective medium for music learning.

Spatial audio has been shown to play a pivotal role in creating compelling VR experiences, yet most VR research up until recently favoured the visual channel (Çamcı and Hamilton, 2020; Broderick et al., 2018; Potter et al., 2022; Kailas and Tiwari, 2021). Recent work has begun highlighting this imbalance by exploring how audio-first design can deepen immersion in VR (Çamcı and Hamilton, 2020). Meanwhile, the influence of embodiment and visual cues on musical performance has been investigated by Yu et al. (2023), who found that high embodiment improved lesson completion rates, whereas strong visual cues enhanced rhythm accuracy. Their results further revealed that low embodiment combined with strong visual cues reduced playing errors, suggesting a trade-off in choosing between these design parameters. Critically, a high level of visual cues and high embodiment together lowered cognitive load and boosted instructional efficiency, highlighting the need for careful orchestration of sensory inputs.

Given the potential of cross-modal design (see Section 6.2.3), VR applications stand to benefit from employing both spatial audio and visual cues optimally. As noted by Chalmers et al. (2009), creating balanced, high-fidelity virtual environments involves trade-offs in design choices, suggesting that carefully tuned combinations of visual, auditory, and interactive elements can elevate VR's capacity to teach music concepts effectively.

This study hypothesises that personalised pitch-colour mappings will provide increased support to the ExpGr in the music test, leading to lower error scores and shorter completion times compared to the pre-defined mapping used by the CtrlGr. Both approaches draw from existing research, raising the question of whether personalised or more structured mappings provide stronger perceptual and cognitive

[7]Virtuoso, `https://virtuoso-vr.com`, accessed 10 August 2024.

[8]PatchWorld, `https://patchxr.com/`, accessed 15 January 2025.

[9]VRTuos, `https://vrtuos.eu/#about`, accessed 10 August 2024.

[10]PianoVision, `https://www.pianovision.com/`, accessed 10 August 2024.

[11]Perfect Pitch, `https://www.meta.com/en-gb/experiences/perfect-pitch/6534981779885606/`, accessed 26 January 2025.

support during tasks. The personalised pitch-colour associations draw on research into synaesthesia and suggest that chromaesthesia is on a spectrum, not a binary conclusion (Campen, 2010). Additionally, engaging with a cross-modal VR design may strengthen these sensory connections, aiding participants in making meaningful pitch-colour associations within the C Major Scale. The pre-defined pitch-colour mappings are based on research in the general population and tend to have less precision. It follows the understandings presented in Section 6.2.1, with lightness and brightness associated with pitch. The colour mapping to the C Major Scale presented to the CtrlGr follows the order shown in Figure 6.4, with the *C4* note mapped to *blue* on the low end of the spectrum, continuing to the high end of the spectrum, with the colour *cyan*. With the pre-defined pitch-colour mappings, CtrlGr participants did not get to form any personal relationship or have extra engagement with pitch-colour associations.

The virtual environment was carefully designed to foster cross-modal associations by incorporating prior research into environment design, locomotion, interactions, auditory presentation, and instructions. These considerations informed the methodology, which is presented in the next section.

Table 6.1: Participant Demographics and Group Assignment Information for Study Three

Category	**Details (All Participants, n=30)**
Total Participants	30
Gender Distribution	Male: 63% (19), Female: 33% (10), Non-binary: 3% (1)
VR Experience	70% reported familiarity (21 participants): Minimal: 14, Occasional: 5, Monthly: 1, Weekly: 1
Musical Background	Plays Instrument: 9 No Instrument: 21 Instruments: Piano: 89% (8), Guitar: 56% (5)
Favourite Genres	Classical: 17%, Pop: 17%, Jazz: 13%, Folk: 13%
Formal Training	Yes: 40% (12), No: 60% (18) With Theory: Yes: 58% (7), Unsure: 33% (4), None: 8% (1)

Category	**Experimental Group (n=15)**	**Control Group (n=15)**
Gender	Male: 10, Female: 4, Non-binary: 1	Male: 9, Female: 6
Musical Experience	No Experience: 10 Beginners: 3 Advanced: 2	No Experience: 8 Beginners: 4 Intermediate: 1 Advanced: 1 Expert: 1

6.3 Methodology

This study aims to investigate whether cross-modal pitch-colour associations can support understanding of musical theory based on the C Major Scale. The study included two groups: an Experimental Group (ExpGr) and a Control Group (CtrlGr), each consisting of fifteen participants. Group assignment was based on demographic information outlined in Section 6.3.1. This study's methodology will be discussed under the headings Participants, Apparatus & Stimuli, Experimental Design, Data Collection, and Data Analysis.

6.3.1 Participants

Thirty participants with diverse demographic and experiential backgrounds took part in this study. Table 6.1 provides a breakdown of participants' gender, VR experience, musical background, music preferences and details on participants from each group (experimental and control groups), with group allocation determined by music experience. Overall, 63% identified as male, 33% as female, and 3% as non-binary, with ExpGr participants consisting of ten males, four females, and one non-binary individual, and the CtrlGr participants consisting of nine males and six females.

This diverse participant profile ensured a broad representation of demographic characteristics, musical expertise, and VR experience, which supports the study's validity and generalisability. Ethical approval was obtained from the Social Research Ethics Committee at University College Cork (Log 2024-159), and the study was conducted in accordance with the Declaration of Helsinki.

A post-hoc power analysis conducted using G*Power indicated low statistical power ($1 - \beta \approx 0.25$); as such, the study may have been underpowered to reliably detect effects of this magnitude. This was due to time constraints and is highlighted as a limitation in Section 6.6.

6.3.2 Apparatus & Stimuli

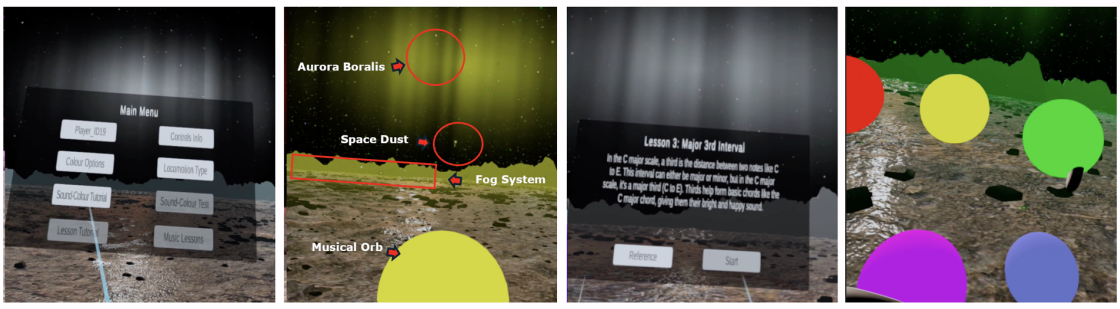


Figure 6.2: Far Left: The main menu with the default environment in the background; **Centre Left:** The environment during pitch-colour associations, showing the aurora borealis, *"space dust"*, fog, and musical orb visual feedback; **Centre Right:** The lesson menu for each lesson, describing the musical sequence under investigation; **Far Right:** The lesson interactions, where participants recreate the audio sequence and receive auditory-visual feedback for each correct interaction.

An `Oculus Quest 2` headset and two `Oculus` controllers were used to run the `Chromaesthetica` VR application. Participants remained seated, each used a swivel

chair to facilitate natural turning, as well as the turning methods used in the application (controller-based and teleportation). `Sony MDR-CD580` on-ear stereo headphones delivered audio playback at a participant-selected volume level.

The study focused on eight musical pitches corresponding to the C Major Scale, ranging from *C4* (261.63 Hz) to *C5* (523.25 Hz). Using `Logic Pro`, each pitch was sampled via a `Steinway Grand Piano` Musical Instrument Digital Interface (MIDI) instrument at a velocity of 100 and exported to a four-second `Audio Interchange File Format (AIFF)` audio file (24-bit, 48 kHz). These sounds were spatialised in `Unity` using the `Audio Spatialiser Software Development Kit (SDK)`. Other audio design elements featured a background *"space"* track, also derived from the C Major Scale, that provided environmental ambience. The footstep sound effects were designed to react to participants' free movement in 3D space and were scripted to reflect the environment and surface realistically. Each of the four sound effects presented slight variations to sound more natural, providing another cross-modal effect in the environment (as highlighted in Section 6.2.2). Text-to-speech `MPEG-1 Audio Layer 3 (MP3)` files were used for auditory instructions at key points throughout the study.

A colour palette used in a study by Palmer et al. (2013) provided a broad range of saturated, muted, light, dark, and achromatic hues (35 colour choices). These colours were applied in `Unity` using *"unlit"* `Universal Render Pipeline (URP)` shaders, ensuring that ambient lighting did not alter their appearance. The palette featured in the colour reference menu, the orb objects participants interacted with, the *"space dust"* particle system, the aurora borealis animation, and the pitch-colour choice reference displayed before each lesson (see Figure 6.2). This consistent implementation of unlit materials preserved the intended colours across different lighting scenarios, ensuring a reliable cross-modal experience for all participants.

6.3.3 Experimental Design

Before the study, a pre-test questionnaire, administered online, gathered demographic information along with participants' musical and VR experience, which helped determine group allocation. Figure 6.3 illustrates the study flow for both groups, with green arrows signifying the CtrlGr pathway and blue arrows signifying the ExpGr pathway.

Upon arrival, each participant was seated in a soundproofed recording studio to minimise external noise. Before putting on the VR headset, they received a brief introduction to the VR HMD controls and the overall study objectives. After donning the headset, participants encountered a *"Start Screen"* displaying their User ID and a *"Group Option"* drop-down menu. The researcher directed them to select the appropriate group (CtrlGr or ExpGr), enabling access to the *"Main Menu"*.

ExpGr participants then entered the *"Sound-Colour (S-C) Tutorial"*—a preliminary practice stage where they explored how to assign colours to musical notes using eight white orbs, each randomly assigned a note from the C Major Scale. Participants experimented with adjusting orb colours and environmental effects (e.g., *"aurora borealis"*, *"space dust"*, `Unity's` fog system), but these practice mappings were not saved. The purpose of the *"S-C Tutorial"* was to familiarise participants with the interaction mechanics and colour-mapping process needed for forming their personalised mappings in the next stage. Once comfortable with these mappings,

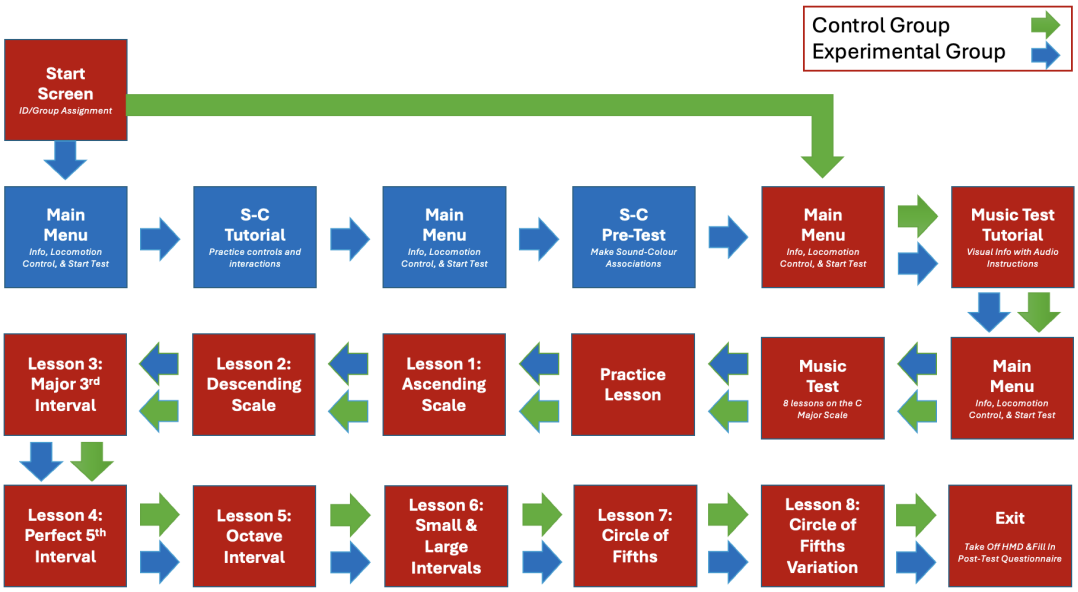


Figure 6.3: Study Three design flow, beginning with the Start Screen. The blue arrows represent the Experimental Group, who completed additional stages to create their own personalised pitch-colour mappings before starting the music lessons. The green arrows represent the Control Group, who skipped these stages and used a pre-defined mapping instead. Both groups completed the same music lessons, but with different pitch-colour mappings.

the ExpGr completed a *"S-C Pre-Test"* (the same process as the tutorial), with an option to confirm or revise any associations before finalising their personalised pitch-colour choices. In contrast, CtrlGr participants skipped the *"S-C Tutorial"* and the *"S-C Pre-Test"*, relying on a predefined pitch-colour mapping supplied by the application (see Figure 6.4). The CtrlGr did not complete the *"S-C Pre-Test"* or tutorial because their pitch-colour mappings were already provided. Having them create personalised mappings, only to then use the pre-defined set during the lessons, could have created a bias and confused the participant by introducing conflicting pitch-colour associations.

After the *"S-C Pre-Test"* for the ExpGr (or directly after the *"Start Menu"* for the CtrlGr), participants were directed to the *"Music Test Tutorial"*, which introduced them to the task involved in the *"Music Test"* section. The interactions and tasks done by the ExpGr in the *"S-C Tutorial"*, were different to the task involved in the *"Music Test"*, with just the grab and release interactions similar but with different auditory and visual feedback associated. The *"S-C"* stages therefore served only to let the ExpGr create their personalised mappings, not to rehearse the music-lesson content. After the tutorial and on entry to the *"Music Test"*, they first undertook a practice lesson to acclimatise themselves to the environment and the lesson mechanics, which included hearing an ascending audio sequence, observing mapped colour cues, and then arranging colour-coded orbs in the correct sequence. With each music note using 3D sound, participants also received an additional auditory cue, as the sound originated from the direction of the corresponding colour-coded orb. Each of the eight lessons that followed focused on a particular aspect of the C Major Scale, progressing from ascending and descending scales to intervals (major third, perfect fifth, and full octave) and then to melodic lines derived from the circle of fifths (see Figure 6.4).

As the lessons started, participants heard the lessons' audio sequence that simultaneously displayed colour changes (based on either their personalised mappings in the ExpGr or the fixed scheme in the CtrlGr). Once the audio ended, they were presented with a brief *"distraction"* countdown featuring rapidly shifting colours. Participants then attempted to reconstruct the target sequence by placing colour-coded orbs onto connector spheres in the correct order, using the mapped colours as their only guide. Any incorrect placement played an error sound and incremented the participants' error score, while correct placements played the associated music note along with the related colour via the environment effects. The next lesson was unlocked once the participant correctly placed all eight orbs.

Upon completing the final lesson, participants received feedback on the individual lesson error scores and the total number of errors. Next, participants exited the VR application, removed the HMD, and completed a post-test questionnaire on the researchers' computer. This questionnaire captured subjective feedback on the VR experience, including ease of use, perceived learning, and any challenges faced during the tasks. In total, this design enabled a direct comparison of learning outcomes and user experiences between individuals who formed their own cross-modal pitch-colour associations (ExpGr) and those who employed a predefined mapping (CtrlGr), thereby assessing the potential impact of personalised cross-modal design on music theory learning.

6.3.4 Data Collection

Data were collected at three stages: pre-test, during-test, and post-test. For the pre-test phase, participants completed a `Microsoft Forms` questionnaire before arriving at the study, providing demographic information, musical background (instrument played, years of experience), and VR familiarity. These responses guided the assignment of participants to either the CtrlGr or ExpGr, aiming for a balanced distribution of musical expertise.

During the experiment, data were captured by the VR application itself and saved locally on the `Oculus Quest 2` as `JSON` files named by participant ID (e.g., Player_ID01). These files recorded timestamped interactions, pre-test pitch-colour associations (for the ExpGr), and all pitch-colour selections made during the music lessons. For each lesson, the application also recorded error scores and completion times. Error scores were calculated by incrementing a value each time a music orb was placed incorrectly in the lesson sequence. Both individual lesson scores and overall totals were captured for each participant.

Immediately after the VR session, participants completed a post-test questionnaire via `Microsoft Forms`. This survey posed 10 to 13 questions (depending on group), including four closed-ended and nine open-ended items (see Appendix C.5). Responses to the open-ended questions contributed qualitative data for subsequent analysis.

6.3.5 Data Analysis

Quantitative Analysis

The quantitative data analysis was conducted using `SPSS` (version 29). The analysis aimed to understand the differences between the CtrlGr and ExpGr participants in

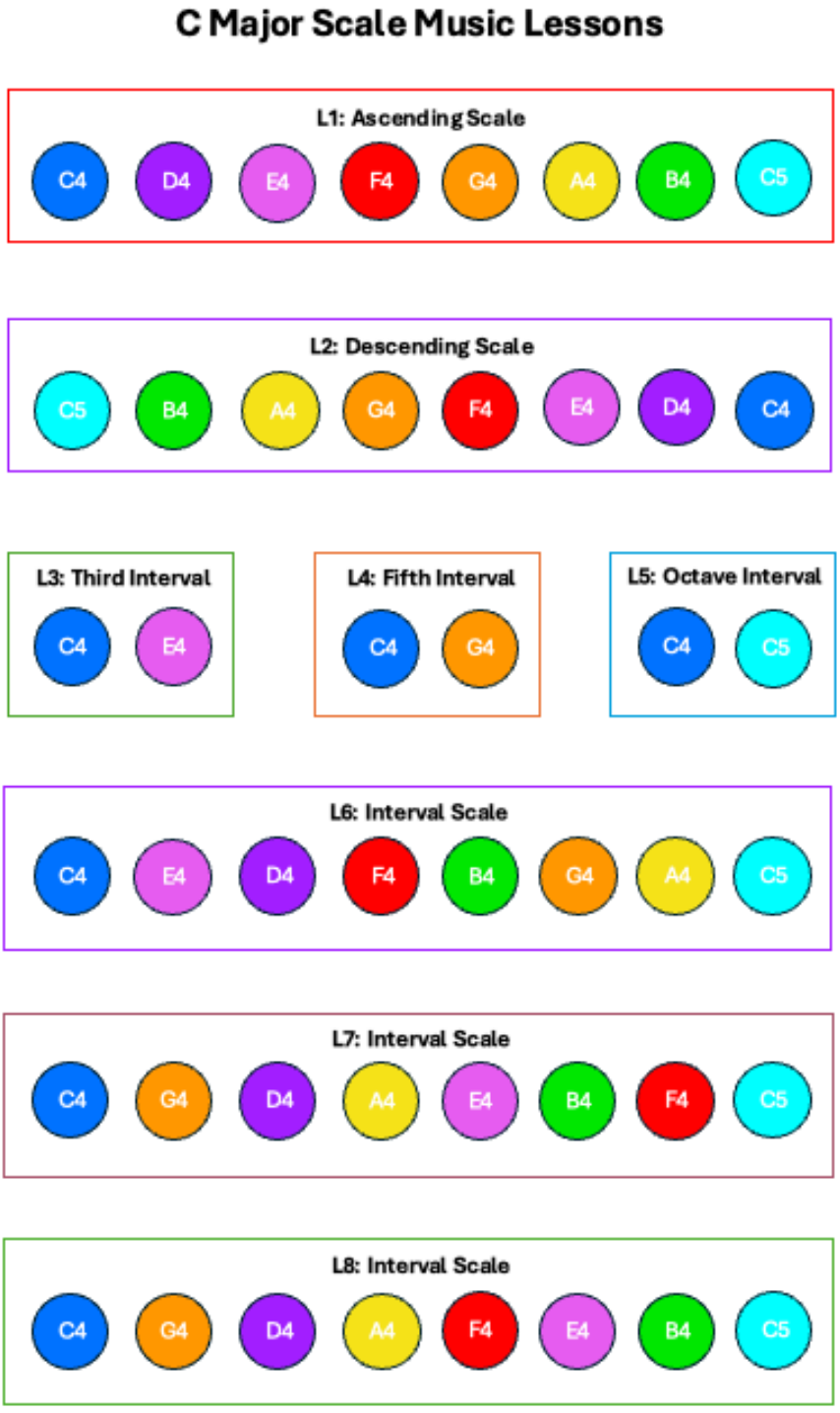


Figure 6.4: The eight music lessons used in the study, showing the note sequences for each lesson and the corresponding pitch-colour mappings used by the Control Group. Both groups completed the same lessons; however, the colours shown here reflect only the pre-defined mappings used by the Control Group. The Experimental Group used personalised pitch-colour mappings, which varied by participant and are therefore not represented in this figure.

their performance of the musical test, and the progress of learning across all eight lessons. Also, the personalised pitch-colour associations made by the ExpGr were analysed to identify patterns of colour choices between the participants. Assumptions of normality and homogeneity were checked, and non-parametric methods were applied when they were found to be violated. The `Mann-Whitney U` test for non-parametric data was used to compare the total error score and total completion times between the control and experimental groups, with the results pertaining to *RQ1*, stated in Section 6.1.

A repeated measures `ANOVA` was conducted to analyse differences in performance across the eight lessons, as well as to test for interactions between lessons and group type, with these results informing *RQ2*. Post-hoc pairwise comparisons were conducted to identify specific lesson-to-lesson differences. For the pitch-colour associations made by the ExpGr (see *RQ3* in Section 6.1), descriptive statistics and visualisations were generated to summarise the frequency of colour choices.

Qualitative Analysis

The questionnaire data taken pre- and post-test were combined into a single `CSV` file by participant ID to support both demographic insights and qualitative analysis. Open-ended responses were analysed using `NVivo 12` following a template analysis approach, which is a form of thematic analysis that offers flexibility to adapt themes and codes through the analysis process (Braun and Clarke, 2019).

Initial *"a priori"* codes (i.e., codes defined before analysis) were developed based on the main areas of interest: *"Sound-Colour Tasks"*, *"Learning Outcomes"*, *"VR Experience, Game Controls and Interactions"*, *"Feedback"*, *"Strengths and Weaknesses"*, and *"Understanding the C Major Scale"*. As the coding progressed, these codes went through two iterations on the data. Some codes were merged (e.g., VR Experience and Game Controls and Interactions), and others contributed to more than one final theme where they overlapped in meaning. This process resulted in the final template themes of *"Sound-Colour Associations"*, *"Learning Outcomes"*, and *"User Experience in VR"*, which better captured participants' perspectives and the broader insights relevant to the study (see Appendix C.6).

6.4 Results

6.4.1 Quantitative Analysis Results

The total error scores and total time taken by each group were analysed. The results from the `Mann-Whitney U` test (see Table 6.2) showed no significant difference in total scores between the control group ($M_r = 14.53$, Median = 15) and the experimental group ($M_r = 16.47$, Median = 18), $U = 98.00$, $Z = -0.602$, $p = 0.547$, $r = 0.11$. Similarly, no significant difference was found between the total time taken in the control group ($M_r = 17.93$, Median = 1059) and the experimental group ($M_r = 13.07$, Median = 952), $U = 149.00$, $Z = 1.514$, $p = 0.130$, $r = 0.28$. These results indicate that the type of pitch-colour mapping used did not significantly affect participants' total scores or total time taken, and do not support the hypothesis that the experimental group would outperform the control group.

Table 6.2: Results from the `Mann-Whitney U` tests comparing the *Error Scores* and *Time Taken* between the Control Group (CtrlGr) and Experimental Group (ExpGr). The table presents the median scores, mean ranks, test statistics, associated p-values, and calculated effect sizes r. These results provide insight into group differences in performance and efficiency during the music learning tasks.

Measure	Statistic	CtrlGr	ExpGr	Test Results
Error Scores	Mean Rank (M_r)	14.53	16.47	$U = 98.00$
	Median	15	18	$Z = -0.60$, $p = .547$, $r = 0.11$
Time Taken	Mean Rank (M_r)	17.93	13.07	$U = 149.00$
	Median	1059	952	$Z = 1.51$, $p = .130$, $r = 0.28$

A repeated measures `ANOVA` was conducted to examine the effect of lessons on performance (error scores and time taken) across eight lessons, with group (control vs. experimental) as a between-subjects factor. When analysing the error scores, `Mauchly's Test of Sphericity` indicated a violation of the sphericity assumption for the main effect of lessons ($W = 0.021, \chi^2(27) = 98.67, p < .001$). Therefore, the `Greenhouse-Geisser` correction ($\epsilon = 0.533$) was applied to the degrees of freedom for the within-subjects effects. The results revealed a significant main effect of lessons on performance, $F(3.73, 104.56) = 11.86, p < .001, \eta_p^2 = 0.298$, indicating a significant variation in performance (error scores) across lessons (see Figures 6.5 and 6.6).

The interaction between lessons and group was not significant, $F(3.73, 104.56) = 0.729, p = .565$, suggesting that the control and experimental groups did not differ significantly in their learning trends across lessons. Between-subjects effects from the repeated measures `ANOVA` revealed no significant differences in performance between the control and experimental groups, $F(1, 28) = 0.052, p = .822$. Tests of within-subject contrasts indicated significant linear and quadratic trends across lessons, $F(1, 28) = 8.54, p = .007$, and $F(1, 28) = 13.22, p = .001$, respectively, suggesting steady improvements in performance with some levelling off at later lessons. Bonferroni-corrected pairwise tests identified several significant contrasts in error scores between specific lessons, which are summarised in Table 6.3.

A repeated measures `ANOVA` was conducted to examine the effect of lessons and group (control vs. experimental) on total time. `Mauchly's Test of Sphericity` indicated that the assumption of sphericity was violated for the main effect of lessons ($W = 0.012, \chi^2(27) = 113.463, p < .001$). Therefore, the `Greenhouse-Geisser` correction ($\epsilon = 0.496$) was applied. The results revealed a significant main effect of lessons on total time, $F(3.47, 97.24) = 28.903, p < .001, \eta_p^2 = 0.508$, indicating significant variation in completion times across lessons (see Figures 6.5 and 6.6).

The interaction between lessons and group was not significant, $F(3.47, 97.24) = 0.350, p = .818, \eta_p^2 = 0.012$, suggesting that both groups exhibited similar trends in total time across lessons. There was no significant difference in overall total time between the control and experimental groups, $F(1, 28) = 0.009, p = .924, \eta_p^2 = 0.000$. A number of lesson-to-lesson comparisons showed significant changes in completion

Table 6.3: Significant Bonferroni-corrected pairwise comparisons for error scores. L = Lesson in the Comparison column. 95% CI shown as [lower, upper]. "Direction" indicates which lesson had higher errors.

Comparison	Mean Diff	95% CI	p (Bonf.)	Direction
L2 – L5	3.567	[0.087, 7.046]	.040	higher at L2
L3 – L6	−3.533	[-6.482, -0.585]	.008	higher at L6
L3 – L7	−3.633	[-6.756, -0.510]	.011	higher at L7
L3 – L8	−2.467	[-4.859, -0.074]	.038	higher at L8
L4 – L6	−4.167	[-6.810, -1.523]	< .001	higher at L6
L4 – L7	−4.267	[-7.288, -1.246]	.001	higher at L7
L4 – L8	−3.100	[-5.138, -1.062]	< .001	higher at L8
L5 – L6	−4.733	[-7.190, -2.277]	< .001	higher at L6
L5 – L7	−4.833	[-7.791, -1.876]	< .001	higher at L7
L5 – L8	−3.667	[-5.704, -1.629]	< .001	higher at L8

time, with full details reported in Table 6.4. For the SPSS output on the within- and between-subject for both the error scores and the time taken, please see Appendix C.8.

Table 6.4: Significant Bonferroni-corrected pairwise comparisons for completion time. L = Lesson in the Comparison column. 95% CI shown as [lower, upper]. "Direction" indicates which lesson took longer.

Comparison	Mean Diff	95% CI	p (Bonf.)	Direction
L1 – L3	34.400	[19.381, 49.419]	< .001	L1 slower
L1 – L4	34.967	[16.270, 53.663]	< .001	L1 slower
L1 – L5	44.467	[33.177, 55.757]	< .001	L1 slower
L1 – L7	−21.967	[-43.579, -0.354]	.043	L7 slower
L2 – L3	48.367	[20.574, 76.160]	< .001	L2 slower
L2 – L4	48.933	[24.419, 73.447]	< .001	L2 slower
L2 – L5	58.433	[30.983, 85.884]	< .001	L2 slower
L3 – L6	−64.500	[-103.209, -25.791]	< .001	L6 slower
L3 – L7	−56.367	[-81.915, -30.818]	< .001	L7 slower
L3 – L8	−45.467	[-69.221, -21.713]	< .001	L8 slower
L4 – L6	−65.067	[-91.126, -39.007]	< .001	L6 slower
L4 – L7	−56.933	[-82.101, -31.766]	< .001	L7 slower
L4 – L8	−46.033	[-68.604, -23.462]	< .001	L8 slower
L5 – L6	−74.567	[-111.668, -37.466]	< .001	L6 slower
L5 – L7	−66.433	[-91.682, -41.184]	< .001	L7 slower
L5 – L8	−55.533	[-80.076, -30.991]	< .001	L8 slower

To explore the pitch-colour associations made by the ExpGr, data were analysed for common patterns, focusing on specific colour codes, colour families (e.g., reds, blues, greens), and brightness levels (e.g., darker or lighter). For individual colours, Figure 6.7 illustrates the frequency of pitch-colour associations for each note in the C Major Scale. *Saturated yellow* was the most frequently chosen colour (10 times overall), with seven of these linked to *B4* (3 times) and *C5* (4 times). *Saturated red*



Figure 6.5: Mean error scores for the experimental and control groups by lesson

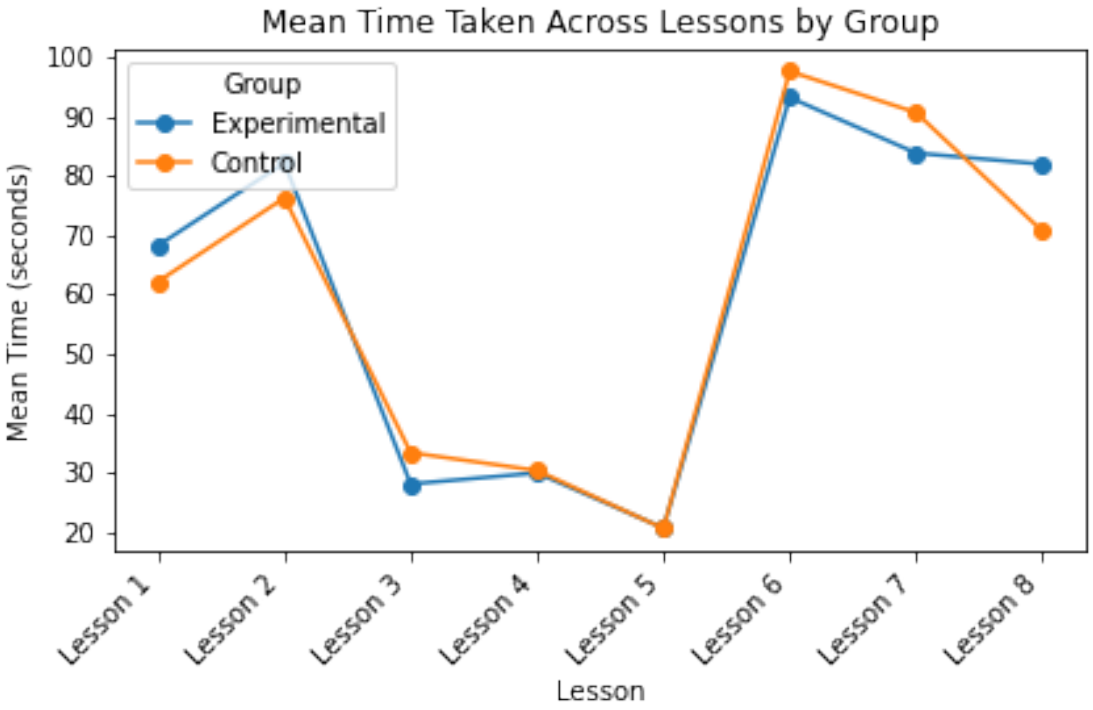


Figure 6.6: Mean time taken (in seconds) for the experimental and control groups by lesson

was split between the lower and upper ends of the scale, including two associations with *C5* and two with *E4*. *Saturated orange* was notably linked to *B4* (3 times). Other frequent associations included *A4* with *light chartreuse* (3 times), *C5* with *light yellow* (3 times), and *E4* with *saturated blue* (3 times).

Colour families are shown in Figure 6.8, which reveals that *yellow* was most frequently associated with higher notes, appearing seven times for *C5* and five times for *B4*. *Chartreuse* was chosen seven times for *A4*, while *red* and *blue* were more commonly associated with lower notes like *C4* and *E4*, each appearing five times. Additionally, lower notes (*C4*, *D4*, *E4*) frequently featured *blue*, *purple*, *red*, and *achromatic* groups, each occurring three times.

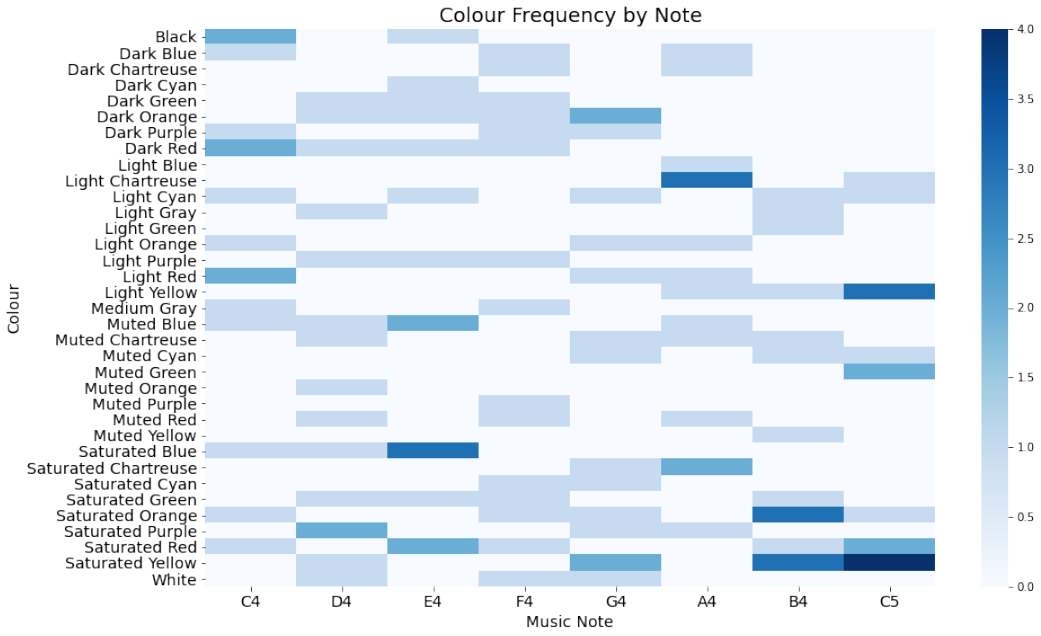


Figure 6.7: This heat map shows the frequency of colour choices for each note in the C Major scale (*C4*-*C5*). Participants selected from a palette of thirty-five colours, which included achromatic, dark, muted, light, and saturated options.

For brightness levels, Figure 6.9 shows a preference for brighter colours (saturated and light) with higher notes, while darker colours (muted and dark) were less frequently associated with lower notes. The two highest notes, *C5* and *B4*, were associated with brighter colours twelve times each, compared to only three darker associations for each. Similarly, *G4* and *A4* were linked to brighter colours ten times, with only five darker associations.

These results suggest a strong tendency to associate brighter colours with higher-pitched notes, though darker colours did not show a consistent pattern with lower-pitched notes.

6.4.2 Qualitative Analysis Results

This section presents the results of the template analysis conducted on the open-ended responses from the post-test questionnaire. The following abbreviations are used to reference participant numbers and groups: Px-ExpGr (Participant x from the experimental group) and Px-CtrlGr (Participant x from the control group). All

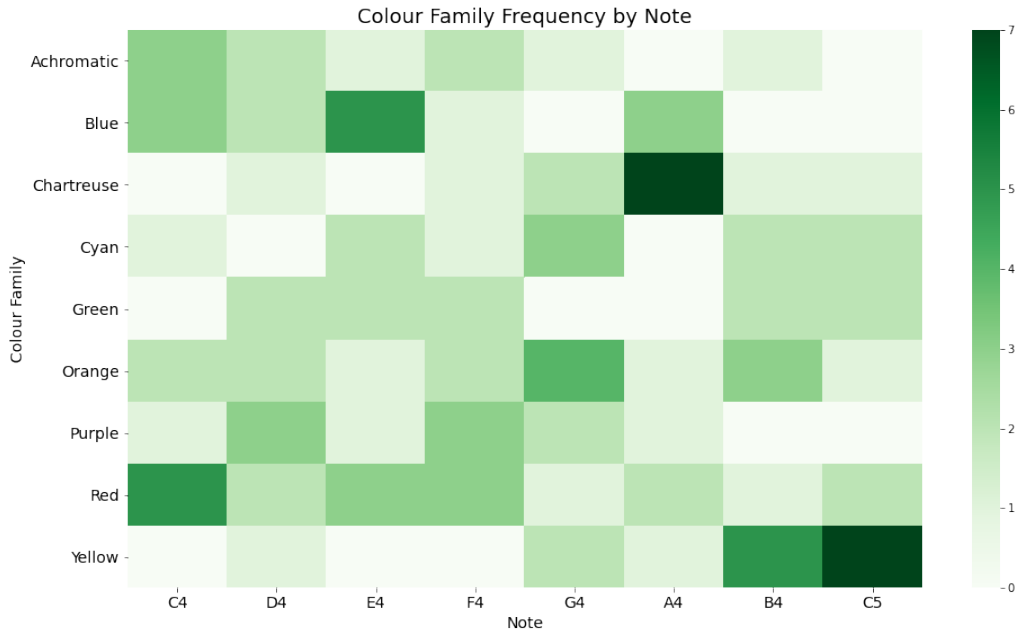


Figure 6.8: This heat map shows the frequency of colour family choices (i.e. reds, yellows, blues, etc) for each note in the C Major Scale (*C4-C5*), with the achromatic, dark, muted, light, and saturated colour options grouped into their respective colour family.

participant quotes are presented verbatim with small edits in square brackets to provide clarity.

Three main themes emerged, reflecting participants' experiences with the VR application: *"Sound-Colour Associations"*, *"Learning Outcomes"*, and *"Usability of the VR Application"*. The following sections discuss each theme in detail, with representative participant quotes provided as evidence.

Sound-Colour Associations

This theme focuses on participants in the ExpGr and their experience forming personalised pitch-colour associations. Three questions from the post-test questionnaire were only asked to the ExpGr; however, several insights from the CtrlGr were added, as they highlighted features related to this theme unprompted. This theme investigates how participants formed associations between sound and colour in the VR environment. It is presented in three subthemes: challenge making sound-colour associations, pre-task impact on the music test, and effect of environment on sound-colour associations.

Challenge making sound-colour associations: The first subtheme within the sound-colour associations theme centres on the challenges participants experienced when choosing colours to match the musical notes they heard. While the ExpGr had more focused questions on these challenges when making their personalised associations, insights were observed across the pre-test and the music test environments in both groups (CtrlGr and ExpGr).

When asked if they found it challenging to associate musical notes with colours in the pre-test environment, P1-ExpGr explained: *"Slightly, yes. Certain notes in isolation reminded me of a colour, and then when played in sequence they ended up*

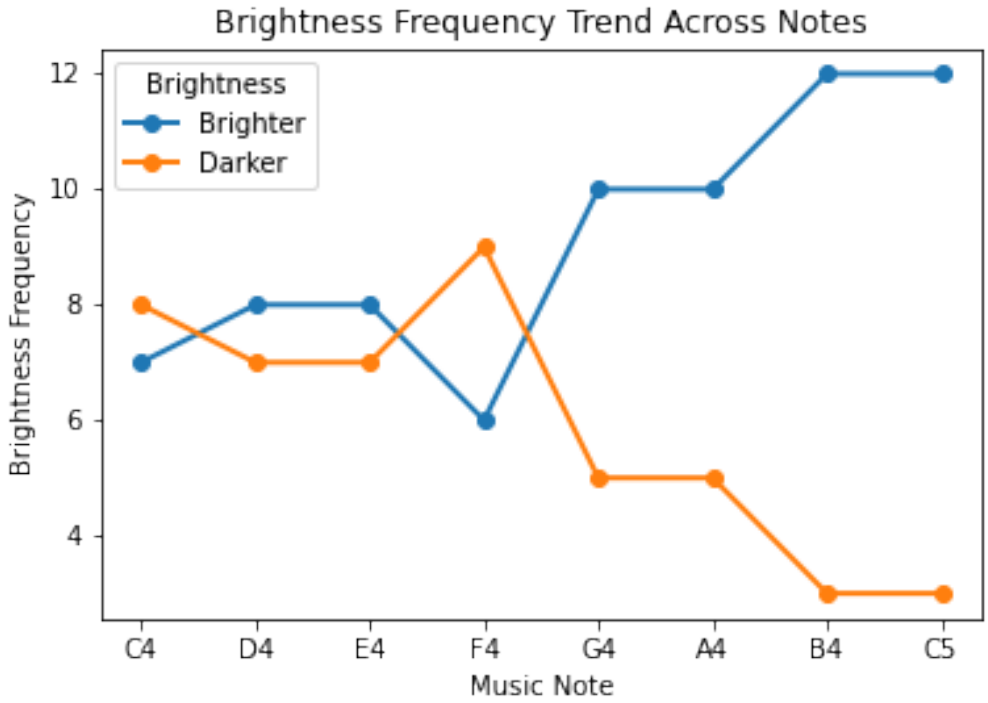


Figure 6.9: In terms of brightness, this graph shows the frequency of grouped colour choices by brightness for each note, with muted and dark colour types classed as *"Darker"*, and light and saturated colour types classed as *"Brighter"*.

sounding "wrong". So I had to choose the colours knowing that I was going to be hearing them in sequence. Often I forgot what part of the sequence I was listening to, which potentially could have changed how I felt a colour affected a note. I associate low notes with dark colours, and high notes with lighter ones, but again, out of sequence it was difficult for me to 100% make that distinction". P15-ExpGr shared similar insights regarding hearing the notes in sequence, saying: *"Yes I did as I was not sure what note would come next. If I could listen to them all in a row then my choice of colour would have changed"*. These responses suggest that pitch-colour associations may not be entirely fixed or one-to-one but can shift depending on the context in which a note is heard. In both cases, participants indicated that hearing notes in isolation led to different colour choices than when those same notes were heard in sequence. This highlights how temporal and relational factors, such as the direction of pitch movement or the expected position of a note within a scale, may influence how a participant perceives and associates colour with pitch. P1-ExpGr particularly emphasised that their colour choices would vary depending on whether the notes were heard in isolation or as part of a melody.

When examining how participants approached the task of associating musical notes with colours, several different strategies were reported. Reflecting on their process, P6-ExpGr remarked: *"I tried to associate higher notes with lighter colours as best I could"*. Meanwhile, P28-ExpGr described a more intuitive approach, stating: *"just didn't thought [think] too much about it and trusted my intuition"*. Others found the task more difficult. For instance, P20-ExpGr stated: *"Sometimes - I found it difficult when two notes were similar as I felt inclined to assign similar colours"*. These varying approaches highlight individual differences in how participants tack-

led the task. Some used strategies consistent with broader population trends, such as associating higher-pitched notes with lighter colours, while others relied on intuition. On investigation of the pitch-colour associations made by participants using a more intuitive approach, they also tended to associate lighter colours, such as *"light yellow"*, *"saturated yellow"* and *"light chartreuse"*, with higher-pitched notes.

Another challenge participants encountered involved the controls for the task. When attempting to save their colour choices, P12-ExpGr commented: *"Sometimes the colour selection was too sensitive, so when A [controller button] was selected the colour changed due to hand movement"*. Similarly, P29-ExpGr remarked: *"Not so difficult, but a little bit challenging to use the controller to select the colour"*. These observations underscore how technical aspects, such as the sensitivity of the controls and participants' familiarity with VR, could impact their ability to make effective pitch-colour associations.

From the CtrlGr's insights that mention pitch-colour associations, a different, more positive experience is expressed. When discussing the challenges of learning musical theory, P14-CtrlGr stated *"I always found musical theory quite daunting and complicated. Associating notes with colours makes it a lot more accessible which I liked"*. P21-CtrlGr mentioned *"I had never previously associated colour with music—but now I can see the benefit. At this point, I am associating some specific notes with specific colours. Blue and Purple seem to be ones that I can recall distinctly"*. With the CtrlGr participants having a pre-defined pitch-colour mapping and not taking part in the pre-test, this could explain the more positive outlook on the pitch-colour associations.

In summary, participants faced difficulties when hearing notes in sequence versus isolation, with some relying on intuition and others associating higher notes with lighter colours. Technical issues, such as controller sensitivity, further complicated the task for some participants. While CtrlGr participants reported fewer issues with pitch-colour associations overall, they did not create personalised mappings and instead engaged with these associations only within the music test context.

Pre-task impact on the music test: This subtheme examines whether the experimental group's pre-test task influenced their performance in the music test. When asked about the impact of the pre-test, P11-ExpGr noted that it was: *"Good practice, improved my performance in the test by getting more used to the different notes in isolation"*. Similarly, P28-ExpGr reflected: *"probably yes. Having assigned the colours on my own probably is closer to what I feel when I hear a certain sound"*. P20-ExpGr added: *"Yes, I felt I was "rehearsed" in which colours to assign as I had spent some time mulling over it already. From a user interface perspective, I also found it much easier to navigate the virtual environment, having practiced"*. These insights suggest that the pre-test positively affected participants' subsequent performance in the music test. The benefits included greater familiarity with notes in isolation, developing personal associations, and having time to rehearse the associations.

One notable advantage of the pre-test, as highlighted by participants in the experimental group, was becoming familiar with the controls. P15-ExpGr commented: *"Yes I think it gave me a slight advantage/help as I had a small bit of practice beforehand and was a little more familiar with the mechanics of the game"*. Similarly, P3-ExpGr observed: *"I don't think so but maybe it gave me more time to practice*

the interactions and get familiar with the changing environment". These responses indicate that spending time acclimatising to the VR environment during the pre-test helped participants feel more confident and capable during the music test.

To summarise, the pre-test task helped participants in the experimental group perform more confidently in the music test by familiarising them with notes, allowing rehearsal, and providing an opportunity to practise using the VR environment. However, the benefits varied across individuals.

Effect of environment on sound-colour associations: The environment in which the participants were immersed during the study was an important aspect of the experience. Many participants from both groups responded positively, describing the environment as engaging, conducive to focus, and giving them a sense of ease. Again, while one question sought insights on just the ExpGr's experience in the pre-test environment, the CtrlGr provided insights on the environment unprompted in other questions. P29-ExpGr remarked: *"It is a very engaging environment and definitely holds the attention. Initially, it was difficult and a bit distracting to get used to the various controls, but once familiar, I found it very helpful"*. Similarly, P4-CtrlGr shared *"it is always awkard [awkward] but I felt that the environment programmed in VR setting was helpful to feel at ease with it"*. P30-ExpGr also praised the immersive qualities, stating: *"It was very nice, the 3D environment felt very real. It was user friendly and the controllers moved easily between the orbs and the colour selection panel"*. The visual feedback in the environment was appreciated by P13-CtrlGr, who stated *" The use of colours and a space-like environment was very cool. I liked looking around"*.

However, opinions concerning movement within the environment were mixed. P15-ExpGr expressed a dislike for the locomotion aspect, noting: *"Personally, I would prefer an environment that required less walking around for this type of exercise"*. In contrast, P1-ExpGr appreciated the physical engagement, stating: *"The environment was simple and visually interesting, the orbs were spaced apart from each other but that was interesting because it made me physically move around the space"*. P7-CtrlGr highlighted that the environment did not cause negative effects, and mentioned *" It was very good and immersive. The good thing about this environment is that it does not make anxious or sick"*.

The immersive environment was positively received, with participants finding it engaging and conducive to focus. Conflicting views on movement in VR were noted, as some appreciated the physical engagement while others found it unnecessary for the task. With the CtrlGr not engaging in movement but just turning via the methods mentioned in Section 6.3.2, a more positive experience by some participants was had.

Theme one summary: Overall, the theme of sound-colour associations revealed a range of participant experiences. While the pre-test task benefited the experimental group in familiarising themselves with the notes and VR environment, challenges were noted, particularly in associating notes with colours when hearing them in sequence after making isolated pitch-colour associations in the pre-test. The immersive environment supported engagement, although issues such as locomotion and control sensitivity posed obstacles for some participants.

Learning Outcomes

The Learning Outcomes theme explores participants' experiences with the VR application, focusing on how they used pitch-colour associations to understand the musical concepts presented. It examines differences in perceived learning between the control and experimental groups and identifies aspects that benefited understanding music theory. This theme also considers the perceived difficulty of the lessons and participants' progress through them, their awareness of musical intervals and their related colours, the design features that supported learning, and participants' views on VR as a learning tool.

Group differences in learning: There appeared to be differences in how participants approached the task and the contrast of learning strategies employed when taking the music lessons. However, little variation was found in the overall learning outcomes when comparing the control and experimental groups, as shown in Figure 6.5 and 6.6. Similar insights were shared across both groups. For example, P6-ExpGr, a participant with no prior music experience, stated: *"I had no experience coming in and i [I] came out of it have an understanding of the basics"*. A similar response was given by P25-CtrlGr, a participant from the control group who also had no music experience: *"Yes. I had no previous knowledge and this made the topic approachable and interesting"*. This indicates that participants in both groups, regardless of prior experience, were able to achieve foundational learning outcomes.

Differences between the control and experimental groups became more evident when considering the role of pitch-colour associations during the music test. Some participants in the control group noted challenges with the pre-defined colour mappings, as the colours did not align with their intuitive associations. P22-CtrlGr explained: *"It was more like a memory test for myself as the colours didn't fully align with sounds for me. I did enjoy the VR aspect of grabbing and placing the elements in specific locations"*. Similarly, P27-CtrlGr described having to adapt to the pre-set mappings, stating: *"I needed to adjust the colour scale because I got used to starting at red for middle C. this became easier toward the end of the lesson"*. In contrast, participants in the experimental group, who could assign their own colours, highlighted the benefits of creating personal pitch-colour mappings. P16-ExpGr noted: *"Positives- fun, interactive and the way they place the sounds with the colours at the start was helpful"*. P28-ExpGr similarly reflected: *"Having assigned the colours on my own probably is closer to what I feel when I hear a certain sound"*. These responses suggest that self-created mappings positively impacted learning, making the lessons more intuitive and engaging for the experimental group.

Differences in learning outcomes between the control and experimental groups were primarily related to pitch-colour associations. Participants in the control group often struggled with the pre-set mappings, as they did not align with their intuitive associations. In contrast, participants in the experimental group found creating their own mappings beneficial, allowing for more personal and meaningful connections with the material. This underscores the importance of tailoring the learning experience to the participants' level of prior knowledge.

Benefit for learning music: Participants overwhelmingly viewed the Chromaesthetica application—and VR in general—as an effective tool for learning music.

They frequently highlighted the pitch-colour cross-modal design, combined with movement and interaction, as particularly beneficial for understanding musical theory. For instance, P8-ExpGr stated that *"Association of musical sounds with colour is very beneficial for music-relevant studies"*, and P27-CtrlGr remarked on the flexibility this design offers, and said, *"I learned to be more flexible with which colours can be associated with music notes"*. P9-CtrlGr emphasised the multisensory nature of the experience: *"I think the use of colour and movement will help people learning music theory as it involves more of the senses and muscle memory, as opposed to traditional teaching"*.

An additional insight pertains to participants' prior musical experience. Two individuals with intermediate or advanced backgrounds (one from each group) reported minimal new learning from the lessons. P13-CtrlGr, for example, explained: *"No, but that is only because I have studied the scale for years prior. The linking of colours to the notes of the scale is very interesting and I can see the benefit of the same for younger or less experienced musicians"*. Similarly, P28-ExpGr, who had 13 years of piano lessons, noted: *"Not that much. Since I had piano lessons for 13 years, I'm quite familiar with the basic harmonies within the scale"*, however did note that *"I enjoyed it, it helps me having another sense involved in the learning process which can be quite supporting.*

These observations suggest that participants with little or no musical training derived the greatest benefit, particularly in grasping foundational concepts like the C Major Scale and pitch-colour relationships. By contrast, those with advanced knowledge already understood these concepts, reducing the novelty and perceived learning gains. The findings indicate that cross-modal associations in a VR context can support musical understanding by engaging multiple senses. Yet, the degree of impact may depend on learners' prior expertise.

Lesson difficulty and progress: Examining the difficulty of the music lessons and participants' progression through them revealed similar feedback from both groups. While participants enjoyed the exercises and found the pitch-colour learning mechanism beneficial, many reported that the difficulty level increased significantly in the later lessons, leading to struggles. P3-CtrlGr, a participant with no prior musical training, remarked: *"It was definitely helpful for me to associate colours with note. I think for me there was a significant jump in terms of difficulty at level 5 — I made 10 mistakes there. Then, I figured out and I was better. This is coming from somebody with zero musical training or skills"*. Similar challenges were noted by P20-ExpGr, who commented: *"I enjoyed the exercise, but found the music theory in the later lessons difficult to grasp. I felt that for most of the notes, I matched colours with orbs based on visual memory (memorising the pattern of the colours shown in the sky in the virtual environment, rather than "feeling" which colour went with which note)"*. These responses suggest that while participants initially found the lessons engaging and manageable, the significant jump in difficulty in the later lessons caused issues, particularly for those with limited prior exposure to music theory.

The use of colours to associate with music notes supported participants in understanding the sequences of notes and lesson content. P14-CtrlGr reflected: *"Yes. Being able to see the scale in colour spectrum made the progressions much easier to visualise and understand"*. P5-CtrlGr echoed this sentiment and emphasised the

importance of the lesson explanations, stating: *"Yes, the explanations were interesting and the colours helped understand the progressions"*. These insights highlight the combined value of the auditory-visual explanations and the colour associations in aiding participants' understanding of the C Major Scale and its progressions.

Participants generally found the music lessons engaging and valuable, with the pitch-colour associations helping them visualise and understand the progressions within the C Major Scale. However, the difficulty of the lessons increased significantly in the final stages, with many participants struggling to grasp the advanced music theory concepts introduced. The auditory-visual explanations and the use of colour mappings were reported as effective tools for task performance, particularly in the earlier lessons. These findings suggest that while the learning mechanism was beneficial, the progression of lesson difficulty could be better balanced to avoid steep jumps that may hinder task performance.

Awareness of musical intervals and colour relationships: This subtheme explores participants' awareness of the musical intervals in the C Major Scale after completing the music test. Participants with no or limited musical experience generally reported gaining a good understanding of the associated musical theory. P2-ExpGr reflected on their learning, stating: *"Yes, the exercise explained intervals such as thirds, fifths and octaves well, and associating each gap with a certain colour helps me hear this better"*. Similarly, two participants with no prior musical training shared that they developed a better understanding of the concepts presented. P12-ExpGr noted: *"Yes, I understood harmonics and variations more because I could reference the sound after the explanation"*. P11-ExpGr also remarked: *"I started with no knowledge of this and definitely began to learn and hear the concepts of intervals"*.

The colour visuals supporting the audio sequences were frequently cited as beneficial for understanding intervals, alongside the spatial positioning of orbs and lesson instructions. Participants with little to no musical experience reported that these elements improved their grasp of interval relationships in the C Major Scale. The combination of auditory and visual cues, particularly colour associations, helped reinforce harmonic concepts and variations.

Application design for learning: Participants provided positive feedback on the application design and how it fostered music learning in VR. The pitch-colour cross-modal approach to the music lessons was frequently highlighted as a key strength. P17-CtrlGr described this positively when asked about the benefits of the music test, stating: *"It's so [such] a fun way of learning and supporting music with colours which is always amazing idea cause it helps deaf people and for children"*. Similarly, P14-CtrlGr emphasised the accessibility of associating notes with colours while also appreciating the virtual environment and narration, saying: *"I always found musical theory quite daunting and complicated. Associating notes with colours makes it a lot more accessible which I liked. I really enjoyed the space/other planet setting, rather than just a room. I would have liked if the narration matched the text at the start of lessons"*.

However, some participants struggled with the test, particularly those with no musical background or limited VR experience. For example, P23-CtrlGr noted: *"What I can say is [it] can be a bit difficult for a person who has no background in*

music. that person may need more practice sessions". P12-ExpGr shared a similar perspective, commenting: *"Somewhat difficult for a non-musically trained person to deal with all the new musical concepts. There were a lot of instructions to deal with"*.

Participants generally appreciated the application design and the cross-modal approach, particularly its ability to make musical theory more accessible through the use of colour associations. However, the effectiveness of the application varied depending on participants' prior musical knowledge and VR familiarity. Those with no musical background found the test challenging, particularly in managing the volume of new concepts and instructions presented. Additionally, participants highlighted factors such as musical experience, VR expectations, and the design of the virtual environment as important elements influencing their learning experience.

VR as a learning tool: Using VR as a learning tool elicited a range of interesting insights, both positive and negative. Participants frequently mentioned immersion, interaction, and engagement as key benefits of using a VR headset for musical learning. Several participants highlighted the advantages of VR over traditional learning methods, such as reading from a book. P3-ExpGr commented: *"It was kind of cool and fun. It was more engaging than learning from a book"*. Similarly, P14-CtrlGr stated: *"I enjoyed it. The active participation felt more fun than just reading from a book"*.

The immersive qualities of the VR environment were another aspect that participants appreciated. P7-CtrlGr reflected: *"It was very good and immersive. The good thing about this environment is that it does not make [me] anxious or sick"*. Despite the benefits, using VR head-mounted displays (HMDs) for learning is relatively new and introduces its own challenges. Depending on the HMD used, comfort and usability issues can arise, which are often highly dependent on the individual. Most participants reported no significant issues, but some highlighted discomfort during prolonged use. P6-ExpGr noted: *"At the start was quite manageable but as the lessons progressed, I began to experience headaches which lowered my ability to listen"*. Similarly, P4-CtrlGr mentioned initial discomfort when adjusting to the new environment but noted that it improved over time, stating: *"It is always awkward but I felt that the environment programmed in VR setting was helpful to feel at ease with it"*.

Participants found VR to be an engaging and immersive tool for learning musical concepts, often preferring it over traditional methods like reading from a book. However, challenges such as discomfort and headaches during extended use were noted, with individual differences playing a significant role. While most participants adapted well to the VR environment, these findings highlight the importance of addressing user comfort in VR-based educational design.

Theme Two Summary The theme of learning outcomes highlights how participants engaged with the application to understand musical concepts, particularly the C Major Scale. Participants with no or limited musical experience reported improvement in their understanding of the musical concepts presented, supported by the use of pitch-colour associations and auditory-visual explanations. While most found the lessons engaging and accessible, the increased difficulty in later stages posed challenges, especially for those with minimal music training. Differences between

the control and experimental groups were observed, with experimental participants benefiting from self-created colour mappings. Additionally, the VR environment was recognised as an immersive and interactive tool, though issues such as discomfort and motion sickness affected some users. These findings underscore the potential of VR for music education, while highlighting areas for improvement in lesson progression and user comfort.

Usability of VR Application

The Usability of the VR Application theme examines participants' experiences using the system, highlighting both positive aspects and suggestions for improvement. It is presented through three subthemes: immersion and engagement; issues with the VR application or device; and usability of the interface. Participants discussed how the virtual environment, interaction, navigation, interface elements, and instructions shaped their experience and offered insights to improve the application for future learning contexts.

Immersion and engagement: One key aspect of the VR application highlighted by participants was its immersive nature and how it enhanced their focus during the task. P6-ExpGr noted: *"I really enjoyed how the VR set immersed me in the lesson and can [got] my full application"*. P7-CtrlGr shared a similar view, commenting: *"VR applications are immersive and help you to stay focused"*.

The interaction mechanics in both the pre-test and music test were also positively received. Participants appreciated being able to manipulate objects directly, as it provided a sense of control and engagement. P1-ExpGr explained: *"Being able to manipulate objects is always good for a user, because it allows a sense of control and impact"*. Similarly, P12-ExpGr highlighted the practicality of the interactions, saying: *"The test were [was] practical being able to drag notes into the correct position on the scale"*. The simplicity and intuitiveness of the interactions were also praised, with P19-CtrlGr stating: *"virtual interaction with the objects, easy to use, very intuitive"*.

Another benefit noted by participants was the gamification of the learning task. P29-ExpGr remarked: *"Gamifying the experience keeps the engagement more fun and adding the competition element also helps"*. These insights demonstrate that features such as immersive environments, intuitive interactions, and gamified elements contributed to participants' engagement and focus, underscoring the potential of VR applications as a viable medium to create learning experiences.

Issues with the VR application or device: When discussing issues related to the VR headset or the immersive VR experience, there appeared to be two main challenges: discomfort caused by extended use of the headset and motion sickness. The time spent on the VR application was a concern for some participants due to the weight and discomfort of the headset. When asked about the challenges with the VR application, P31-CtrlGr commented: *"[VR is] suitable for the short term session, without a longer experience it would be hard to say"*. Reinforcing this, P6-ExpGr stated: *"The longer it went on I started to experience headaches"*. While most participants did not report headaches, other discomforts, such as the noticeable weight of the headset, were distracting. P20-ExpGr explained: *"The weight of the*

headset was noticeable at times. — From the halfway point of the exercise, it was quite noticeable and sometimes distracted me a little. Nothing severe though, no headaches etc as a result".

The second issue raised in the analysis was motion sickness, which is common when using VR HMDs. Although only a few participants experienced this, it was noted as a challenge. P1-ExpGr stated: *"The motion sickness aspect I mentioned previously, but there is very little that can be done about that due to each individual having a unique reaction to VR"*. Locomotion within the VR environment was identified as a potential cause of motion sickness. P15-ExpGr reflected: *"I find walking around in a VR environment a bit nauseating so it would not be my preferred method of learning. However, if you could develop the same lesson with less walking and looking around, that would be great"*. Another participant, P18-ExpGr, reported dizziness when interacting with the orbs, stating: *"I got dizzy when I placed the orbs and the screen shook a little bit"*.

To summarise, participants highlighted two main challenges with the VR application: discomfort from extended use of the headset and motion sickness. The weight of the headset and the duration of the session were noted as sources of distraction for some users. Motion sickness, though less common, was linked to locomotion and specific interactions within the VR environment. While these issues are largely individual and common in VR applications, minimising discomfort through careful design and selecting a suitable headset could improve user experience.

Usability of the interface: Participants highlighted several points about the interface's usability, particularly regarding the audio-visual presentation of instructions. A common issue was the inconsistency between the text and audio descriptions, which distracted and confused some users. P14-CtrlGr suggested: *"Maybe instead of the voice-over and textboxes providing different information, they could be the same"*, Similarly, P9-CtrlGr remarked: *"Matching audio and text would have assisted me"*. The length and complexity of the instructions were also noted, with P3-ExpGr commenting: *"Maybe the instructions and guides could be more streamlined and more challenging lessons"*.

Another design suggestion focused on the audio sequence playback in the music test. Several participants expressed the need to easily reference individual notes or lesson sequences. P12-ExpGr suggested: *"Being able to click an orb during the test to get its note, or get the note reference easily again, would be helpful"*. Similarly, P25-CtrlGr mentioned: *"It might be helpful when you click on an orb to hear it before placing it, as I had to rely on my memory of the colours rather than their sound"*. Other participants shared similar thoughts. P8-ExpGr noted: *"Remembering the sounds associated with colours it [was] challenging for me"*. P21-CtrlGr added: *"Longer colour and note display times- the ability to freely "play" the notes at will with the colours prior to the training"*.

In summary, participants raised two key points regarding the interface's usability: synchronisation between audio and visual instructions and the need for more accessible audio feedback during the music test. Discrepancies between text and voice-over instructions created confusion for some, while the limited ability to replay individual notes and lesson sequences posed challenges in associating sounds with colours. While these design choices aligned with the study's objectives, they provide valuable insights for improving the application's usability and accessibility

in the future.

Theme Three Summary On this theme of Usability of VR Application, participants praised the immersive VR environment for enhancing focus and engagement, with intuitive interactions like manipulating orbs benefiting task performance. However, challenges such as motion sickness, headset discomfort, and limited audio replay in the music test were noted. Discrepancies between audio and visual instructions also caused confusion. Suggestions for gamification and improved audio feedback highlighted areas for refining the application to enhance usability and support learning experiences.

To conclude this section, the results of the template analysis were presented under the three main themes of *"Sound-Colour Associations"*, *"Learning Outcomes"*, and *Usability of VR Application"*, with several subthemes attached to each theme. The findings will be discussed in the next section along with the quantitative results, framed around the research questions presented in Section 6.1.

6.5 Discussion

The following discussion will integrate the quantitative and qualitative analysis results under each research question presented in Section 6.1. Interesting insights were found on how participants associate colours with the musical notes, the role of environment, locomotion and interaction when conducting the music-based task using the VR HMD.

6.5.1 Research Question One

> *Q1: In an immersive VR cross-modal music test, how does participants' performance differ between the control and experimental groups?*

The quantitative analysis revealed no significant difference in total error scores or total time taken between the control (CtrlGr) and experimental (ExpGr) groups, suggesting that both personalised and pre-defined pitch-colour mappings provided a comparable learning experience. This fails to reject the null hypothesis, stating that the ExpGr would outperform the control group in the music test. While the ExpGr had the extra *"S-C Tutorial"* and *"S-C Pre-Test"* tasks—getting to familiarise themselves with the controls, environment, and practice and make personalised associations—this showed no effect on performance compared with the CtrlGr.

Despite failing to reject the null hypothesis, the ExpGr demonstrated a slightly lower mean time taken, likely due to their pre-test task, which helped familiarise them with the controls and environment (see Section 6.4.2). Conversely, the ExpGr had slightly higher error scores, potentially due to challenges in creating their pitch-colour associations during the pre-test. As highlighted in Section 6.4.2, some participants experienced controller sensitivity issues that inadvertently altered their colour choices. Additionally, participants who selected visually similar colours for different notes may have struggled during the music lessons, despite the design safeguards in place to minimise such errors (e.g., unique colour assignments and the option to revise choices).

Although the Experimental Group was expected to show improved learning outcomes compared to the Control Group, `Mann-Whitney U` test results for both the error scores and total completion times revealed no significant differences. This suggests that individual differences in musical experience, pitch-colour associations, and learning styles may have influenced the outcomes. Some participants in the CtrlGr expressed dissatisfaction with the pre-defined mappings, noting that some colours did not align with their intuitive associations (e.g., P27-CtrlGr, who associated C with red). This suggests that personalised colour mapping features may offer greater flexibility, allowing users to align the system with their own cognitive associations, potentially improving engagement and learning efficacy. Participants from both groups (e.g., P27-CtrlGr and P6-ExpGr) expressed a preference for features available in the other condition, which highlights differences in how individuals perceive and associate sound and colour. This would align with the view that synaesthesia, or sound-colour relationships, may exist on a spectrum (Campen, 2010).

Moreover, some participants followed established cross-modal trends, associating higher-pitched notes with brighter colours (Spence, 2011; Martino and Marks, 1999), while others, like P28-ExpGr, adopted more intuitive approaches, which interestingly still follow the cross-modal trends previously mentioned. This highlights that these established cross-modal associations (higher-pitch to brighter colours) are prevalent, even on a subconscious level. Musically experienced participants (e.g., P13-CtrlGr and P28-ExpGr) leveraged prior knowledge to aid in the lessons, further demonstrating the diversity of strategies. These findings underscore the variability in pitch-colour associations and the importance of accommodating individual differences when designing cross-modal learning applications.

While this study primarily focused on pitch-colour relationships, other auditory features, such as melody, key, and timbre, remain unexplored. Future research could investigate how these features interact with cross-modal design to optimise learning outcomes in music education.

6.5.2 Research Question Two

> *Q2: Does participant "performance"—defined by error scores, completion times, and reported learning experiences—improve over the eight lessons for both groups?*

The repeated-measures `ANOVA` and post-hoc analyses revealed significant performance improvements across the eight lessons, attributed largely to the structured lesson design. Lessons One and Two involved playing all eight notes of the scale (*C4-C5*), while Lessons Three to Five focused on simpler intervals (3rd, perfect 5th, and octave), requiring only two correct notes at a time. Lessons Six to Eight reintroduced all eight notes in more complex patterns (e.g., the circle of fifths). This tiered structure explains the variation in error scores and completion times, with participants performing differently when transitioning between full scales, simpler intervals, and more complex sequences.

Error scores showed a clear trend of improvement across the lessons, with mean errors decreasing from Lesson One to Lesson Five. A spike in errors in Lesson Six (mean score increasing from 0.87 in Lesson Six to 5.60 in Lesson Eight) aligns with the increased complexity and number of interactions required. However, errors decreased again in Lessons Seven and Eight, indicating progressive adaptation

and learning. These findings align with Merchant et al. (2014), who observed performance gains in virtual environments for Kindergarten to twelfth Grade (K-12) education.

When comparing the two groups, the experimental group exhibited higher mean error scores in six out of eight lessons, with the control group having a higher mean error score in Lessons One and Seven. This suggests that while the experimental group had additional time in VR making personalised pitch-colour associations, it did not translate to a significantly better performance. As noted in Section 6.5.1, challenges such as similar colour choices and mistakes in the pre-test could have impacted the experimental groups' performance.

Qualitative insights from the template analysis reinforce these findings, highlighting both the benefits and challenges of pre-defined and personalised mappings in the music-based task. Both groups found the pitch-colour associations useful for understanding scales and intervals. However, the sharp increase in error scores in Lesson Six, reflected in the quantitative data, was also noted in participants' feedback (see Section 6.4.2). While the lessons were designed to progress from basic scales to more complex structures, the interval lessons (Lessons Three to Five) were relatively simple, making the transition to Lesson Six feel abrupt. Reorganising the lessons—beginning with simpler intervals, followed by ascending/descending scales, and then introducing the more complex circle of fifths sequences—could have provided a more gradual learning curve. This approach aligns with scaffolding, a method proven to effectively teach music education by breaking down complex material into manageable steps (Wong et al., 2023).

Another insight on the use of visual memory mention in Section 6.4.2 highlights how participants may shift their strategy as task complexity increases. Rather than some participants relying on an intuitive or sensory connection between pitch and colour, they drew more on spatial and visual memory. However, this does not necessarily mean cross-modal associations were unhelpful. Instead, it may suggest that the visual feedback in the environment supported memorisation when cognitive demands increased. It also raises the possibility that cross-modal integration in VR can serve a dual purpose: supporting intuitive associations early on, and then helping to anchor learning through spatial memory when tasks become more difficult. As previous research has shown multisensory training can enhance memory and learning effectiveness (Shams and Seitz, 2008).

Comments from P1-ExpGr and P15-ExpGr (see Section 6.4.2) suggest that perceived pitch-colour associations varied depending on whether notes were heard in isolation or within a melody/sequence. This could have influenced performance across lessons, as different audio contexts may have altered participants' colour choices. While musical features like melody or key can affect cross-modal associations (Bresin, 2005), research suggests that pitch is the primary auditory unit driving sound-colour associations, rather than broader structures like intervals or melodies (Ward et al., 2006).

Overall, both groups followed similar performance trends, seen in Figures 6.5 and 6.6, with insights on the lesson design informing potential changes to future studies. Participants' insights on the perception of colour in different musical contexts highlight the complexity of sound-colour cross-modal associations, with interesting avenues for future exploration.

6.5.3 Research Question Three

Q3: What kinds of pitch-colour associations did participants of the experimental group make?

To explore this question, pitch-colour associations were analysed across three levels: colour code, colour family, and brightness. Results revealed several patterns. At the colour code level, *saturated yellow* was most frequently associated with the two highest notes (*C5* and *B4*), while *saturated orange* and *saturated blue* were each associated with *B4* and *E4* at a frequency of three (see Figure 6.7). These findings align with prior research suggesting that brighter colours are commonly associated with higher-pitched sounds, and darker colours with lower-pitched sounds (Marks, 1987; Spence, 2011).

Similar trends were observed in the colour family analysis (see Figure 6.8). *Yellow* was frequently linked to *C5* and *B4*, *chartreuse* with *A4*, and darker colours like *blue* and *red* with the lowest notes (*E4* and *C4*, respectively). These associations also support the tendency for darker colours to correspond to lower-pitched notes and lighter colours to higher-pitched notes (Marks, 1987; Spence, 2011).

While the experimental group's associations were personalised and most likely different from the control group's pre-defined mapping (which followed a full saturation colour spectrum from blue for *C4* to cyan for *C5*), the experimental group still demonstrated a preference for darker colours at the lower end and brighter colours at the higher end of the scale. This reflects findings in the literature that pitch-brightness associations are common in both synaesthetes and the general population (Ward et al., 2006). A point of note, the study in this paper did not test participants for synaesthesia. The brightness-level analysis further confirmed this tendency, with brighter colours (saturated and light classifications) consistently associated with higher-pitched notes (see Figure 6.9). Some participants also expressed this preference during the pre-test, as noted in Section 6.4.2.

Despite these general trends, no universal mapping emerged. While there were similarities with the control groups' mapping (e.g., *blue-C4* and *yellow-higher notes*), no participant in the experimental group exactly replicated the pre-defined mappings. This highlights the highly individual nature of pitch-colour associations, influenced by personal relationships with pitch and colour stimuli.

The findings for this section suggest that future designs could benefit from a flexible approach, providing a default mapping based on pitch-lightness relationships while allowing users to customise mappings to reflect their unique associations. It is also recommended to consider other sound-colour associations and their potential impact on pitch-colour connections. Such adaptability could accommodate both general tendencies and individual preferences, enhancing the efficacy of cross-modal music education tools.

6.5.4 Research Question Four

Q4: How can immersive virtual environments for musical education improve going forward?

The findings from both quantitative and qualitative analyses highlight the potential of immersive VR environments that use cross-modal design to support learning

music theory. Participants demonstrated consistent learning progress across the lessons (see Section 6.5.2) and found features such as the virtual environment, interaction design, and instructional guidance to be engaging and immersive. However, some challenges, including previous musical experience, motion sickness, and minor discomfort with the VR headset, were also reported.

The minimalist *"space"* environment was designed to reduce visual clutter and enhance focus (Gkoumas and Shimi, 2022). It also supported pitch-colour associations through *"space dust"* and an aurora borealis (see Section 6.4.2), reinforcing immersion and cross-modal connections. This aligns with research on designing virtual environments that promote engagement and elicit affective states conducive to learning-related processes (Mantovani, 2001; Riva et al., 2007).

Some participants (e.g., P22-CtrlGr) briefly mentioned spatial location as beneficial to understand the lessons. In the music test, 3D audio played music notes from their associated colour-coded orbs, though participants did not explicitly acknowledge this auditory feature designed to help locate the orbs. Research suggests spatial location associates with pitch, with higher notes perceived as higher and lower notes as lower (Curwen, 2018). While height-based spatial mapping was considered, it was excluded to maintain focus on colour associations; however, this could be explored in future studies.

The interaction design was well-received, with participants finding the pre-test and music lessons intuitive and engaging. In the pre-test, users grabbed orbs to trigger auditory stimuli and adjust colours, while the music lessons involved connecting orbs to different musical connectors to recreate a melody line. These interactions fostered a sense of control and engagement, aligning with research showing that effective VR interactions can enhance performance and support learning outcomes (McMahan et al., 2006; Kuhail et al., 2022; Merchant et al., 2014; Villena-Taranilla et al., 2022).

Also, visual features like shape and size have been shown to have cross-modal associations with auditory pitch (Spence, 2011; Curwen, 2018; Marks, 1987). Participants indicated that the pitch-colour interactions were compelling, implying that adding manipulable visual features could further enrich engagement and provide valuable cues for understanding complex musical concepts. The study highlights that carefully designed interactions can enhance immersion and support learning experiences while identifying additional elements to support music education in VR.

Clear guidance was essential to participants' experiences, with auditory, visual, and sandbox tutorials helping introduce tasks. Most found the instructions clear, especially when audio and text were combined, though minor misalignments occasionally caused distractions (see Section 6.4.2). Ensuring consistency and streamlining instructions could further reduce cognitive load (Mutlu-Bayraktar et al., 2019). This study aimed to minimise overload through multi- and cross-modal cues and self-paced instruction, though minor issues persisted.

The wide variation in participants' musical backgrounds made designing a universally challenging learning environment difficult. As highlighted in Section 6.4.2, participants with intermediate or advanced musical experience in both groups indicated that their prior knowledge, rather than pitch-colour associations, primarily influenced their performance in the music test. While they acknowledged the application's potential benefits for beginners, they viewed it more as an interesting concept than a learning tool for themselves.

This study primarily targeted beginner learners, though many participants—while not musicians—had some prior music education. Forty per cent (12) had received formal training at some stage, but only seven participants had studied music theory, with four unsure. Introducing multiple difficulty levels tailored to prior experience or incorporating alternative musical systems, such as non-Western frameworks, could provide a more universally new learning experience for all participants.

As mentioned in Section 6.4.2, some participants reported minor discomforts, such as headaches or motion sickness, particularly during the pre-test phase, which required more movement. A study by Boletsis and Cedergren (2019) found that controller-based locomotion increases the likelihood of having motion sickness symptoms compared to teleportation. To address this, the study offered flexible locomotion options, including free movement, teleportation, and turning methods (snap or continuous), with most tasks designed to be stationary.

The results of this study demonstrate that immersive VR environments have substantial potential as an effective tool in supporting musical education. Future improvements should focus on designing simplified yet engaging environments, refining intuitive interactions, ensuring consistency in multimodal instructions, and providing flexible options for locomotion to accommodate user preferences and comfort. By addressing these considerations, VR-based learning tools can be further optimised to deliver engaging, effective, and accessible educational experiences.

6.6 Limitations & Future Work

Reflecting on the study, several potential limitations emerged from participant feedback. The Experimental Group needed to take the extra steps in the *"S-C Tutorial"* and the *"S-C Pre-Test"* to form the personalised pitch-colour associations before undertaking the music lessons. This extra time in the VE could potentially have been advantageous, with the participants getting more familiar with the environment and the interactions. While the music lesson design was not exactly the same as the sound-colour pre-test task, there were similarities. The Experimental Group completed two additional stages prior to the Music Lessons, with the only shared feature being the basic grab-and-release interaction. Unlike the *"S-C Tutorial"* and *"S-C Pre-Test"*—where participants actively adjusted colours and received immediate feedback—the Music Lessons involved no colour manipulation, and auditory-visual feedback was delayed until after each orb was placed. This distinction helped reduce any risk that the extra stages unfairly advantaged the Experimental Group.

Another limitation was the within-lesson performance analysis. Repeating each lesson multiple times could have better captured performance trends and lesson difficulty variations, but at the cost of longer sessions, increasing risks of motion sickness, fatigue, and reduced focus.

Lesson design posed another challenge, particularly in the transition between Lessons Five and Six. Participants struggled with the shift in complexity, suggesting that restructuring the sequence, starting with interval-based lessons (two notes), followed by ascending/descending scales (eight notes), and concluding with the complex melodic patterns, could provide a smoother progression. Alternatively, adding transitional lessons between the interval and more complex lessons may ease this shift, an important consideration for future cross-modal music learning studies.

Another issue stemmed from the auditory-visual sequence playback at the start

of each lesson. Some participants missed crucial information while adjusting to the environment. A countdown or a visual cue could help refocus attention before playback begins. While some participants suggested an option to replay lesson sequences, this would reduce the immediacy of pitch-colour responses central to the study. A compromise could be allowing one additional playback, with its use factored into the final score.

Consistency between auditory and visual instructions was also a concern (see Section 6.5.4). While the instructions were designed to align the auditory and visual instruction stimuli, participants picked up on a few slight misalignments. These misaligned instructions were highlighted as distracting, emphasising the importance of synchronised multimodal cues to minimise cognitive load.

A final limitation of this study, which should be addressed in future research, is the relatively small sample size. Recruiting enough participants for a between-subjects VR study proved challenging due to limited resources. This constraint is reflected in the effect sizes reported in Table 6.2, which show a small effect for error scores and a small-to-medium effect for completion times. Future studies with greater resources and support should aim for at least thirty participants per group to improve statistical power and the generalisability of the findings.

Future research could explore VR cross-modal music learning experiences by varying virtual environments, incorporating additional musical features (melody, timbre, or key), and integrating other visual elements—such as size and shape—with known cross-modal links to pitch. While this study focused on the C Major Scale, expanding to other scales and complex theories could provide deeper insights into cross-modal associations in immersive learning.

Focusing more on personalisation also presents an avenue for further exploration. Incorporating learning styles and personality traits could enhance engagement, with AI offering opportunities for adaptive lesson design. O'Toole et al. (2021)(P1) demonstrated that different personality types and learning styles can influence how participants form pitch-colour associations. AI could refine pitch-colour mappings based on user preferences and continuously tailor content for a more effective, personalised learning experience.

6.7 Conclusion

This study set out to investigate how immersive VR environments, that emphasise auditory-visual cross-modal design, can be used to support music learning, exploring a pre-defined and personalised pitch-colour mapping. The results demonstrate that both groups displayed similar performance trends overall and across the eight lessons, failing to reject the null hypothesis. It also found that when making personalised pitch-colour associations (experimental group), participants tended to associate lighter colours with higher notes.

The `Chromaesthetica` application was well received by participants, with a highlight on the environment design (help with focusing on the task), pitch-colour interactions (easy to use and engaging), and lesson design (intuitive and well described). These findings underscore the potential of immersive technologies to address gaps in traditional music education by offering engaging environments that harness multiple senses to deliver musical theory exercises using an interactive and intuitive interface. In doing so, the study contributes to a growing body of literature on experiential

and multisensory learning, indicating that VR holds promise as a pedagogical tool in music education.

While a full discussion of this study's limitations and potential avenues for future work appears in Section 6.6, respectively, a brief overview underscores the need for further investigation into the lesson design and the potential for the expansion of research into other musical features apart from pitch and how AI can help facilitate cross-modal learning in VR environments. By illustrating the value of cross-modal design in VR-based music education, this study paves the way for innovative, highly interactive learning experiences that can transform how we teach and engage with music.

6.8 Chapter Reflection

The study presented in this chapter builds on the findings from Chapter 5, designing an immersive VE that not only supports pitch-colour cross-modal association tasks, but also applies these associations within a learning context. It addresses the sub-question introduced in Section 1.3: *"How Do Personalised & Pre-Defined Pitch-Colour Cross-Modal Mappings Compare in Supporting Learning Outcomes in VR-Based Music Exercises?"*

This study employed design principles discussed in Section 2.5, focusing on multisensory integration, embodied cognition, and cognitive load theory to create a learning environment that supports engagement and understanding through cross-modal strategies.

Building on Chapter 4, which established the viability of using VR for pitch-colour associations, and Chapter 5, which explored how immersive features influence engagement and usability, this final study integrated user feedback and design insights into a single, cohesive learning experience. Specifically, a calm and focused "space" environment was chosen, informed by the template analysis and quantitative results from the previous chapter, to provide a setting conducive to concentration, engagement, and learning.

The primary outcome of this study was that both participant groups—those using predefined mappings and those generating personalised mappings—showed a similar trend of improvement across the music lessons. Among the personalised group, there was a general tendency to associate higher pitches with lighter colours and lower pitches with darker tones, reflecting patterns found in previous cross-modal research.

Although no significant differences in learning outcomes—in terms of error scores or time taken—were observed between the two mapping types, the study highlights the individuality of participants' pitch-colour associations. The range of personalised mappings revealed variation in perceptual connections, suggesting that while some users naturally align with established trends, others may form unique associations based on internal or experiential factors.

These findings reinforce that pitch-colour cross-modal associations differ across individuals, highlighting the need for further exploration of VR cross-modal design in learning environments that take this diversity into account. While pre-defined mappings offer a consistent association for analysis, the ability to personalise may increase user engagement and help users connect more deeply with the task.

This study provides valuable insights into the central research question: *"How can the cross-modal associations of pitch and colour be enhanced through virtual reality?"* Several key insights emerge from this consideration. Firstly, this study investigated how VR can support the formation and application of pitch-colour associations within a meaningful task, such as learning basic music theory. Secondly, it underscores the variability in how users perceive and interpret sensory connections, influenced by factors such as musical experience, age, or personal preference. Thirdly, it highlights that users engaging in structured tasks within an immersive VR context begin to reflect more deeply on how their senses associate, suggesting the potential for VR to foster both cognitive and perceptual awareness of cross-modal links.

While earlier studies in this thesis investigated the presence and consistency of pitch-colour associations in VR, this study offered a real-world use case—music education—that situates these associations in a meaningful learning experience. Doing so reinforces the potential of immersive environments to enhance cross-modal connections, not simply through visual fidelity or sensory richness but through purposeful, user-centred design.

The following chapter will conclude this investigation, synthesising the insights from Chapters 4, 5, and 6. It will revisit the sub-questions, evaluate the extent to which the central question has been answered, and discuss the limitations of this research while proposing future directions for this evolving and multidisciplinary area of study.

Chapter 7

Conclusions and Future Work

This chapter begins by bringing together the findings from the three studies presented in Chapters 4, 5, and 6. Each chapter addressed a specific research question and together they build a trajectory towards answering the central question of this thesis: *"How can the cross-modal associations of pitch and colour be enhanced through virtual reality?"* Reflecting on the key outcomes from each study, this chapter highlights how the individual findings contribute to a broader understanding of cross-modal perception and its use in immersive VR technologies.

Following the summary of the studies, the discussion focuses on six major themes generated across the research. The first theme (see Section 7.2.1) explores the evolution of the study design throughout the PhD research. An iterative approach guided the research, starting with insights gained from a low-immersion VR setting using a known online consistency test platform. This evolved into more immersive environments that tested specific features, ending with a final VE that incorporated these insights to explore pre-defined and personalised pitch-colour cross-modality in a music learning context. The process was shaped by both practical constraints and the emergent understanding gained at each stage.

The second theme (see Section 7.2.2) examines the role of immersive VR features, specifically how the understanding and implementation of immersive design evolved throughout the research, and how immersive features such as environment design, audio and visual presentation, and interaction mechanics supported the formation of pitch-colour associations.

The third theme (see Section 7.2.3) explores the contrast between personalised and pre-defined sound-colour associations. This raises questions about the commonality of cross-modal mappings and highlights the individual variability in how users perceive and learn from multisensory stimuli. The concept of viewing synaesthesia (especially sound-colour synaesthesia) on a continuum, as previously mentioned in Section 2.2.3, also informs the exploration between pre-defined and personalised pitch-colour associations. This theme reveals implications for designing adaptive learning tools that cater for diverse learning preferences.

The fourth theme (see Section 7.2.4) examines cross-modal integration in greater depth, focusing on how multiple sensory channels interact in VR and how congruent associations influence perception and engagement. This leads to a broader reflection on the cognitive processes underpinning multimodal understanding.

Early research in this PhD highlighted emotion as a factor influencing pitch-colour associations (O'Toole et al., 2019, 2021) (P1). Although explicit emotional

responses became less central as the research progressed, insights from these earlier studies informed the exploration of VR environments as affective spaces capable of shaping cross-modal experiences, which is the topic of the fifth theme (see Section 7.2.5). Research by Riva et al. (2007) positions VR as an affective medium. Building on this knowledge, this thesis incorporates affective design elements, such as environments with calming or positive qualities, to address the central research question, enhance participants' engagement, and strengthen participants' perceptual connections between pitch and colour.

Finally, the sixth theme (see Section 7.2.6) highlights individual differences, showing how people's lived experiences and prior knowledge of music and VR can shape their pitch-colour associations.

After discussing the main themes of the research (see Section 7.2), a summary of the key contributions—encompassing both theoretical advances and practical applications—is provided (see Section 7.3). Building on the presentation of the study findings, main themes, and contributions, the next section brings these elements together informing the central research question (see Section 7.4). To conclude, the limitations (see Section 7.5) of the research are discussed, followed by suggestions for future work (see Section 7.6). The chapter then closes by returning to the central research question—bringing together cumulative insights to evaluate how the thesis advances our understanding of cross-modal associations in immersive virtual environments (see Section 7.7).

7.1 Summary of Findings

This section presents the findings from the research conducted throughout the PhD, culminating in the three studies that form the basis of this thesis. The key outcomes from these studies will be highlighted below.

The research presented here has been informed by multiple disciplines, including HCI, psychology, and music, with early-stage work exploring pitch-colour associations outside VR. Early studies by O'Toole et al. (2019) explored auditory-visual associations using guitar playing within a desktop application, and O'Toole et al. (2021) (P1) examined the role of emotion and other factors as intermediaries between auditory and visual senses. Although not included as thesis chapters, these studies provided important context and direction for the subsequent VR-focused research presented in this thesis.

These initial studies provided valuable insights into user engagement, emotions, personality traits, and participants' understanding of auditory-visual interactions. Although personality traits, musical instruments, and explicit emotional responses became less central as the research evolved, these early findings established foundational insights guiding the shift towards investigating pitch-colour associations within immersive VR contexts. Ultimately, this led to identifying VR features that could effectively support and enhance cross-modal auditory-visual connections.

While previous research has extensively investigated pitch-colour cross-modal associations, and auditory-visual connections more generally, using traditional desktop-based and paper-based methods such as consistency tests (Ward et al., 2006) and speed classification tests (Parise and Spence, 2009), there has been limited exploration of these associations within VR environments. This gap presented a valuable opportunity to investigate how immersive VR features could influence

pitch-colour associations.

The following subsections summarise the findings from each of the three VR studies detailed in Chapters 4, 5, and 6, outlining their contributions towards addressing the thesis's central research question.

7.1.1 Study One

Following the initial research discovery period of the PhD, the exploration of pitch-colour cross-modal associations using VR technologies was an area that was relatively unexplored and thus merited investigation. The first step in this exploration was to conduct a consistency test, comparing a traditional test environment with the same application presented on a VR HMD. This study addressed the following research question presented in Section 1.3: *"Is VR a viable platform for conducting pitch-colour association tests compared to desktop environments?"*.

The results of this study (see Chapter 4) revealed no significant difference in pitch-colour association scores between the two environments. This was the expected outcome, which enabled a foundation for the research where being present in a low-immersion VR space using a HMD without designing for an immersive experience did not affect the score when taking the pitch-colour consistency test. Additionally, the presence of pitch-colour cross-modal associations in VR was confirmed at a similar level to that shown in the desktop version of the test. While these results helped provide a foundation for exploring cross-modal associations in VR, additional insights from the participants' feedback offered further suggestions for using VR.

Promising results from the post-study questionnaire revealed a preference for taking the test in the VR environment, with benefits such as increased focus, better control, engagement, and greater immersion, highlighted by participants. Although the low-immersion VR setting did not significantly change how participants made pitch-colour associations compared to the desktop test, their positive feedback on core VR features suggested that a more immersive virtual environment might better support enhancing cross-modal pitch-colour associations. This study informed the design of the second study, which explored the different immersive features (such as interactions, locomotion, environment and spatial audio) and how participants responded to the changes in environments when interacting with pitch-colour associations.

7.1.2 Study Two

Following the insights from the first study presented in Chapter 4, the second study focused on the research question: *"What role do immersive features play in supporting pitch-colour cross-modal associations in VR learning environments?"*.

Building on the previous study, which confirmed the presence of pitch-colour cross-modality in VR and identified beneficial features of the VR setting (such as focus, control, audio perception, and immersion), this study proposed a deeper exploration of these factors in a more immersive VR environment. This study followed an A/B design exploring two VEs with different features, such as environment setting, locomotion, audio and visual presentation.

A *"forest"* and a *"space"* environment were created to explore two contrasting settings that have been shown to have different potential benefits to participants in

VR. Participants provided positive feedback on both; however, the *"forest"* showed significantly higher SSQ scores compared to the baseline score. Furthermore, the participants highlighted some visual presentation issues, which detracted from the overall experience. For the *"space"* environment, which has a more minimal design, participants found it easier to focus on the task at hand, while also feeling immersed in the *"other worldly"* environment. This finding confirms the qualitative results of the first study, which found that the basic dark background used for the test helped the participant maintain focus.

With the two environments also presenting different locomotion, auditory, and visual presentations, other insights were found. Although each environment used a different locomotion method, participant feedback indicated that locomotion preference was highly individual, with some favouring free movement and others preferring teleportation. Although participants did not specifically comment on the 3D audio presentation of pitch-colour stimuli in the *"forest"*, they did note that the ambient 3D environmental audio contributed to a greater sense of immersion. In contrast, the sparsely used 2D audio in *"space"* received little mention in participant feedback. The *"synaesthetic-like"* visual presentation in the *"forest"* was highlighted as a feature that participants found engaging and helpful when making pitch-colour associations, compared to *"space"*, which did not have this additional visual aid.

This study provided key insights into how to design an immersive and engaging environment that could support the understanding between auditory and visual modalities in terms of pitch-colour associations. Providing balance is key: trying to understand individual differences of the users and presenting an environment that can have a strong impact on the majority of people is challenging, and, in terms of cross-modal association in VR settings, relatively unexplored.

7.1.3 Study Three

The third study presented in Chapter 6 explored a use case for pitch-colour cross-modal associations, with a learning environment focused on basic musical theory. Informed by the results of the second study, a single environment was created to help participants learn different intervals of the C Major scale. Additionally, two different pitch-colour mappings were used in the exercise, with half the participants using pre-defined associations and the other half forming personalised connections to use in the music theory exercises. This design choice explored whether mappings based on previously identified cross-modal associations (higher pitch and brighter colours) (Ward et al., 2006; Spence, 2011) or more personal pitch-colour association, affect learning of music theory.

The results showed no significant difference in error scores and time taken between the two groups, and many insights were gained from the participants about the further development of the environment design and how to present cross-modality in a music learning application. In the personalised group, many participants independently selected colour associations similar to the pre-defined mappings, which were based on previous cross-modal research. Brighter colours like yellows were often linked to higher-pitched notes, while darker colours were associated with lower pitches.

The findings from this study suggest that an adaptive system could be beneficial—starting with a pre-defined mapping based on common associations, and

gradually refining the pitch-colour pairings to align with each user's personal preferences through continued use.

Having summarised the main findings of each study, the following section explores the broader themes that emerged across the research. These recurring themes offer further insight into how immersive VR environments may influence cross-modal associations and help frame the overall direction of the investigation.

7.2 Themes Across Studies

The research presented in this thesis has contributed to building a stronger understanding of pitch-colour cross-modal associations within VR settings. In particular, it has helped demonstrate how to balance key design elements, such as immersion, presence, and interaction, to create an environment that supports stronger perceptual connections. A range of VR features were explored across the studies, with some proving more effective than others in supporting cross-modal connections between pitch and colour.

The following sections highlight the key themes that were generated across the three studies, and how they contributed to answering the central research question. Together, these insights shed light on how VR can immerse users in rich, hyperreal environments that foster deeper engagement with the relationship between auditory and visual perception.

7.2.1 Iterative Design Evolution

The design of each study presented in this thesis was developed iteratively, shaped by the practical challenges of working in immersive VR and the need to understand better how people form cross-modal associations in these environments. Rather than adopting a fixed design, each study was tailored to incorporate insights from the previous study. This approach aligns with best practices in virtual reality development, where user feedback and iterative refinement are essential to ensuring usability, presence, and meaningful engagement (Jerald, 2018).

Iteration allowed for progressive refinement of the design elements, including audio spatialisation, colour selection tools, and locomotion. For example, early feedback on the overwhelming visual density of the *"forest"* environment in Study Two informed the shift to a more minimal and perceptually coherent *"space"* setting in Study Three. Similarly, the evolution of colour input, from an open picker to thirty-five structured choices, responded to the need for clarity and participant agency in forming meaningful associations.

These iterative refinements reflect common principles in VR UX development, where early prototypes are tested and improved through user feedback loops (Soegaard, 2025). Particularly in educational applications involving cross-modal experiences, iteration ensures that interface design supports learning and adapts to users' perceptual and cognitive needs. As Makransky et al. (2019) points out, many companies and institutions still adopt VR for learning simply because it increases presence, without considering whether that presence actually benefits learning outcomes. An iterative approach helps address this by allowing designers to test features, gather user feedback, and adjust the level of immersion to create more meaningful and effective educational experiences.

Overall, the iterative design approach was instrumental in refining the VE to support meaningful pitch-colour associations. The process revealed how even small design choices, such as locomotion or environment adjustments, can impact the focus and engagement required to enhance cross-modal clarity. Future research could benefit from introducing feedback cycles earlier in the design process, especially when creating learning experiences based on subtle perceptual cues and user engagement.

Building upon the iterative design evolution discussed, attention is now turned to specific VR features that have substantial implications for user experience. Examining elements such as environment design, interaction mechanics, and immersion provides clarity on their individual and collective roles in enhancing cross-modal understanding.

7.2.2 Impact of VR Features: Environment, Interaction and Immersion on Pitch-Colour Tasks

Throughout the development of this research, it became clear that the design of specific VR features impacted participants' ability to engage meaningfully with pitch-colour tasks. While these features did not directly determine the clarity of cross-modal associations, they created the perceptual and cognitive conditions under which such associations could emerge more effectively.

Across the studies, environmental design seemed to be an important factor in supporting the formation of pitch-colour associations. Environments that were minimal, visually clear, and free from unnecessary distractions, such as the *"space"* setting, helped users concentrate on the task without being overwhelmed by too much sensory information. These findings align with VR learning research highlighting the importance of reducing extraneous load in immersive contexts to support sustained attention and task relevance (Makransky et al., 2019).

In addition to environmental structure, interaction design played a crucial role in developing relationships between the pitch and colour stimuli. The ability to physically manipulate objects in Studies Two and Three—such as selecting orbs and placing them in sockets—supported active participation and increased users' sense of control. These embodied interactions helped ground the task in intuitive actions and contributed to a more immersive and engaging experience, consistent with the principles of embodied learning (Lindgren and Johnson-Glenberg, 2013; Wilson, 2002).

Spatial audio contributed positively to the immersive quality of the environment, particularly in Study Three, where it aided users in locating colour-matching objects within the virtual space. As previously highlighted, spatial audio can enhance immersion and task performance while also helping to reduce cognitive load (Potter et al., 2022; Huang et al., 2024). Although it was not the primary driver of engagement or focus in this study, it complemented other design features. It added a sense of realism and spatial context that some participants found helpful.

However, not all VR features enhanced the experience. Locomotion control was a recurring issue, with several participants in Studies Two and Three reporting discomfort or disorientation when navigating the environment. These findings align with broader research that identified locomotion as a frequent source of usability and simulator sickness concerns in VR environments (Jerald, 2018). Additionally, sensitivity in controller input and the physical discomfort associated with wearing

a headset for extended periods occasionally disrupted the immersive experience and impacted concentration. In some cases, misalignment between audio cues and on-screen instructions reduced task clarity and contributed to momentary breaks in engagement. Together, these issues show that sensory features alone do not guarantee a positive VR experience unless integrated in ways that align with users' perceptual and cognitive needs.

These findings support the broader principle that VR feature design must be guided by cognitive and perceptual integration, rather than sensory richness alone. As Jerald (2018) notes, human-centred VR systems are most effective when features are designed not in isolation, but as part of a unified user experience. In this research, the interplay of environment, interaction, and auditory-visual feedback—not any single feature—enabled participants to engage with cross-modal associations in a focused and meaningful way. Building on this, the studies also show that individual differences play an important role in how users experience these features.

While environmental and interaction features greatly influenced participants' engagement and clarity of perception, another crucial consideration emerged: the distinction between personalised and pre-defined pitch-colour associations. Exploring this contrast further clarifies how individual variability can inform adaptive VR design strategies.

7.2.3 Personalisation vs Pre-Defined Cross-Modal Associations

One of the focuses of this research was the comparison between personalised and pre-defined pitch-colour mappings. This was informed by previous work in both cross-modal association and synaesthesia literature, which together suggest that these phenomena may exist along a spectrum (see Figure 2.1). Cross-modal associations tend to show common perceptual link—such as higher pitches connecting to brighter colours (Ward et al., 2006; Spence, 2011)—across many people. In contrast, synaesthesia is usually defined by highly individual and stable sensory pairings that remain consistent over time (Curwen, 2018; Farina et al., 2017). For example, a synaesthete might associate the colour yellow with the C4 note, and this association typically remains unchanged throughout their lifetime.

Throughout the three studies presented in this thesis, a trend emerged that aligns with previous findings in the literature: participants frequently associated higher pitches with brighter colour choices, while lower pitches were associated with darker colours. In particular, Study Two (Figure 5.6), and Study Three (Figure 6.9) showed a preference for brighter colours corresponding to higher notes, and darker colours with lower notes, consistent with previous findings (Ward et al., 2006; Spence, 2011). However, while both studies revealed a strong trend for higher pitches being linked with brighter colours, the association between lower pitches and darker colours appeared less consistent in Study Three. One possible explanation relates to the emotional tone of the VR environments. Previous research has shown that emotional responses can influence pitch-colour associations, with lower pitches often linked to sadness and darker colours, and higher pitches to joy and brighter colours (O'Toole et al., 2019, 2021). Additionally, VR environments themselves may subtly shape user mood. It has been found that natural and cosmic environments — such as forests and space—can evoke positive emotional responses in virtual settings

(Riva et al., 2007; Chirico et al., 2018). Although these emotional responses were not directly measured, it is possible that the overall positive tone of the environments contributed to a stronger bias toward brighter colour choices, even for lower-pitched notes.

Importantly, the third study directly compared personalised mappings (allowing individual choices) with pre-defined mappings (based on established cross-modal trends). This comparison provided valuable insights. Personalised mappings enabled participants to engage more intuitively, suggesting that individual preferences could strengthen pitch-colour associations by adding personal meaning and deeper engagement—important considerations in educational contexts. As highlighted in Section 2.5.2, engaging in interactions, such as grabbing the musical orbs and manipulating the colour, can help with memory retention, conceptual understanding, and reducing cognitive load. Conversely, pre-defined mappings offered structured guidance, potentially reducing cognitive load and providing clearer, universally recognisable relationships. This structured approach aligns with CTML, emphasising that coherent and integrated presentation across multiple sensory modalities enhances understanding and recall (Mayer, 2005), supporting the value of structured cross-modal mappings in educational VR settings.

These findings suggest that VR learning environments adopting pitch-colour cross-modal design may benefit from a hybrid approach that integrates both personalised and pre-defined mappings. Personalised mappings promote user engagement and deeper understanding by allowing individuals to construct associations that align with their own perceptual and emotional frameworks. In the personalised group, participants physically interacted with musical orbs and selected colour mappings through direct manipulation—an approach consistent with embodied cognition theory, which posits that learning is enhanced when abstract concepts are grounded in sensorimotor action (Lindgren and Johnson-Glenberg, 2013; Wilson, 2002). At the same time, pre-defined mappings offer structure and clarity, supporting learners who benefit from more guided, universally recognisable relationships between modalities. By combining these approaches, VR learning tools can accommodate diverse cognitive preferences and maximise the accessibility and effectiveness of cross-modal association tasks.

Acknowledging the significance of personalised versus pre-defined associations prompts deeper consideration of the cognitive processes underlying cross-modal integration. Understanding how users integrate auditory and visual stimuli within VR environments enriches the discussion of effective multisensory design.

7.2.4 Cross-Modal Integration

Given the limited research on cross-modal associations within VR, this thesis sought to explore how pitch-colour associations might be enhanced through the immersive features of the technology. While auditory-visual cross-modality has previously been demonstrated in virtual environments (Malpica et al., 2020), no prior studies have focused specifically on the pitch-colour relationship or examined how this association functions within VR.

Across the three studies, participants could form connections between pitch and colour in VR settings. As discussed earlier in relation to personalised and pre-defined mappings, participants often associated higher pitches with brighter colours

and lower pitches with darker tones, this pattern is consistent with previous research outside of VR (Ward et al., 2006; Spence, 2011). This was supported by participant feedback in the qualitative questionnaire findings (see Section 6.4.2), where they described their decision-making processes when linking pitch to colours.

Building on these findings, this research further aimed to explore how understanding of these associations could be enhanced and applied in VR use cases, such as the music exercises developed in Chapter 6.

As highlighted in Chapter 2.5.1, VR's ability to combine multiple sensory modalities offers unique potential for supporting cross-modal learning. Synchronised and semantically congruent stimuli across different senses can strengthen multisensory integration, which has been shown to enhance memory and improve learning outcomes (Alwashmi et al., 2024; Shams and Seitz, 2008).

The design of the VR tasks in this thesis drew on principles from cross-modality, multisensory integration, embodied cognition, and affective design. Interaction design evolved iteratively across the studies, from basic pointing to more refined grab-and-move mechanics, and this was frequently cited by participants as beneficial for both engagement and making intuitive associations between pitch and colour. These richer interactions, with immediate visual feedback linked to auditory input, allowed participants to integrate stimuli more effectively across modalities.

One of the more challenging aspects of the design was presenting colour options in an intuitive and usable way. The colour selections were different in each study and were informed by findings from the previous study. The first study utilised an online synaesthesia testing platform with a full colour picker, but translating this into 3D interaction in VR proved difficult. Recreating the same level of control would have required navigation across the *"x"*, *"y"*, and *"z"* axes to represent hue, saturation, and value, adding unnecessary complexity. To simplify the interaction, the second study limited selection to hue only, but this approach received some negative feedback for being too restrictive.

In response, the third study introduced a colour palette of 35 pre-set options, covering light, dark, muted, and saturated variants. This hybrid approach proved effective, with no participants reporting insufficient choice. The palette was based on Munsell colour values, consistent with those used in previous cross-modal studies (Palmer et al., 2013).

Although this iteration improved usability, there is still room for refinement of the presentation of colour choices. A system combining a default pitch-colour mapping with layered customisation options could cater to novice and experienced users. Kim et al. (2022) found that combining discrete and continuous selection methods in VR colour pickers led to better performance and usability, suggesting that a hybrid approach may be ideal for future applications.

Alongside interaction design and multisensory integration, another important aspect influencing pitch-colour associations is the affective design of the VR environments, which is explored in the following section.

7.2.5 Impact of Affective Environments on Pitch-Colour Associations

Previous research conducted as part of this PhD, but not a direct focus of this thesis, highlighted the role of emotion in how people form pitch-colour cross-modal

associations (O'Toole et al., 2019; Palmer et al., 2013) (P1). However, in Chapters 4, 5, and 6, the focus shifts towards the design of affective VR environments and how they can support participants in forming stronger perceptual connections.

The earlier work mentioned above provided insights that guided the use of affective elements in the VR environments to support cross-modal perception. This approach carried through into the design choices, building on VR's potential as an affective medium for shaping user experiences. This aligns with how VR is seen in the literature—as an affective medium that can shape how people feel based on the design of the virtual space (Riva et al., 2007). In Study Two, two contrasting virtual environments were created: a *"forest"* scene, chosen for its calming and restorative qualities, and a *"space"* setting, which was more abstract and minimal to help focus attention (see Section 5.2.4).

Both of these environments were also influenced by research into awe-inspiring virtual spaces. Chirico et al. (2018) found that natural and cosmic environments—like forests and outer space—effectively evoke positive emotions such as awe and curiosity. These emotions, in turn, can enhance attention and foster mental openness. While awe was not directly measured, the environments in Study Two were designed with distinct emotional tones in mind to help create immersive and engaging experiences.

Participant feedback reflects these affective intentions. The *"forest"* was described as *"relaxing"* and *"natural"*, while the setting of *"space"* was described as *"soothing"* and *"peaceful"*. Although participants did not directly say that the environment's mood affected their pitch-colour choices, many felt the environments helped them focus and stay engaged.

This approach to designing affective VR environments draws on educational research, which shows that emotions can influence learning. Positive emotions, like curiosity and enjoyment, are known to boost motivation and help people stay engaged, while negative emotions can reduce focus and make learning harder (Pekrun and Stephens, 2012). Cognitive load theory also explains that when learners are distracted or emotionally affected, it can overload their mental resources and reduce their ability to learn effectively (Sweller, 2011). Tyng et al. (2017) further shows that emotional states can impact working memory and focus, thus closely linking emotion to learning performance. In this research, the environments were designed to support a positive and focused mindset and help participants feel comfortable and relaxed so that they could better focus on the pitch-colour association tasks. Rather than directly matching specific emotions to individual notes, the goal was to create a positive, calming setting that reduced distractions and supported attention. The studies aimed to create optimal conditions for forming cross-modal associations by placing participants in a calm and immersive space.

Designing for affect can help situate participants in an environment that supports the formation of cross-modal connections in a more intuitive way. Future work should continue to explore how affective design in VR can support learning and perception. By carefully aligning the emotional tone of the environment with the task at hand, more engaging and effective learning experiences can be created. At the same time, affective VR environments may not elicit the same response from everyone. This highlights the importance of developing adaptable systems that can respond to individual needs and preferences.

7.2.6 Individual Differences in Cross-Modal Learning

A recurring theme across all three studies was the variability in how participants formed and interpreted pitch-colour associations. While certain trends—such as higher pitches consistently being linked with brighter colours—were reliably observed, the specifics of these associations often differed between participants. This individual variation fits with existing findings in cross-modal perception and synaesthesia research, which show that while certain perceptual pairings (like pitch-brightness) are commonly shared (Ward et al., 2006), they can also be shaped by personal experience, musical experience, cultural background, or the specific context in which the stimuli are presented (Palmer et al., 2013; Whiteford et al., 2018; O'Toole et al., 2019).

Differences in musical background notably influenced how comfortable participants felt and their approach to making pitch-colour selections. The findings from Study Three also suggest that participants with more musical experience often approached pitch-colour tasks with an awareness of pitch structure and theory, while those with less experience tended to rely more on sensory intuition and the cross-modal cues provided by the VR design. Additionally, some individuals demonstrated an especially acute sensitivity to relationships between their senses, particularly evident among those engaged in creative disciplines such as music, visual art, or design. Research in synaesthesia shows that synaesthetes often gravitate toward artistic professions, with their cross-modal experiences deeply informing their creative processes (Campen, 2010; Ward et al., 2025). This link between how people connect their senses and creativity highlights how challenging it can be to design multisensory learning environments that can be universally adopted.

Participants' prior experience with virtual reality also emerged as a meaningful factor influencing their performance and comfort. In particular, participants with limited or no previous VR exposure occasionally encountered challenges navigating virtual spaces or manipulating virtual objects effectively. This unfamiliarity led to occasional frustration or distraction, especially during tasks involving more complex interactions (such as those introduced in Study Two). Conversely, participants familiar with VR technology typically adapted more quickly, demonstrating higher initial comfort and potentially enhanced engagement. Therefore, varying levels of VR familiarity should be taken into consideration when designing interactive multisensory tasks, highlighting the necessity of accommodating different degrees of prior experience through adaptive interface designs.

An earlier phase of this research, which formed part of my PhD work but is not included as a core study in this thesis, also considered the role of personality in shaping emotional responses to musical stimuli (P1). The study suggested that traits such as *Conscientiousness*, as well as emotional sensitivity, can influence how participants interpret the emotional character of single notes, potentially shaping their pitch-colour associations. This points to a possible connection between individual differences in perception and stable personality traits.

Further evidence for this relationship comes from Chamorro-Premuzic and Furnham (2007), who found that personality and intelligence shape how individuals engage with music. For example, people high in *"Extraversion"* and intelligence may engage with music more analytically or cognitively, while those high in *Neuroticism* and low in *Conscientiousness* are likelier to use music for emotional regulation. These patterns suggest that emotional engagement with auditory stimuli is medi-

ated by personality, and therefore, sensory learning tools—particularly in VR—could benefit from being responsive to such traits.

These findings support the design of adaptive VR systems that can adjust emotional tone, interaction style, or feedback modality in response to personality-driven engagement styles. By aligning system responses with individual users' emotional and perceptual tendencies, such systems could enhance the accessibility and impact of cross-modal learning experiences.

Together, these observations emphasise the importance of developing adaptive VR learning systems capable of responding to diverse user characteristics. A rigid, uniform approach to interface design and sensory mapping is unlikely to accommodate the complexity and range of individual differences observed. Instead, effective cross-modal learning environments should offer flexibility in colour mapping options, interface complexity, interaction modalities, and even the emotional tone of the virtual experience. By providing multiple adaptable elements, VR systems can support a broader range of perceptual, cognitive, and experiential profiles, ultimately enhancing accessibility and effectiveness across diverse user groups.

Having explored the recurrent themes that shaped the studies, the following section consolidates this research's contributions. These contributions not only reflect on theoretical understandings but also deliver practical guidelines for designing immersive, multisensory VR experiences.

7.3 Contributions

This section outlines the key practical and theoretical contributions arising from the research described in this thesis. Drawing on seven published papers and a paper awaiting publication in the `Computers & Education: X Reality Journal`, three of which are presented as core chapters, this work advances knowledge in the area of pitch-colour associations within immersive VR environments.

This research contributes to ongoing conversations about how cross-modal associations can be effectively used in the design of VR tools, especially in relation to music learning and how users engage with pitch and colour. It provides insights into how users interact with multisensory environments, highlighting the importance of usability, engagement, and adaptability in supporting effective learning.

The following subsections summarise the specific contributions under two main categories: practical contributions, which relate to the design and implementation of VR systems, and theoretical contributions, which advance understanding in cross-modal cognition, multisensory learning, and synaesthesia research.

7.3.1 Practical Contributions

Usability and Engagement of Cross-Modal Tasks in VR Music Learning

The practical contributions of this thesis are evident in the design insights it offers for developing VR-based music learning applications. Currently, the use of virtual reality in music education—particularly for music theory—is limited, and no known VR platform explicitly leverages pitch-colour cross-modal associations to explore learning potential. Chapter 6 addressed this gap, demonstrating that carefully designed cross-modal interactions can enhance user engagement and facilitate learning

tasks in immersive environments.

Several interaction design features implemented across the studies showed effectiveness in supporting user engagement and task clarity. Participants consistently highlighted the intuitive nature of object manipulation—specifically, grabbing and placing virtual objects onto connector points—as a critical element in maintaining their active participation and focus. These interactions align closely with existing research indicating that direct 3D manipulation can increase task efficiency in virtual environments (McMahan et al., 2006). Moreover, previous studies on embodied cognition in VR contexts suggest that physically interacting with virtual content not only supports engagement, but may also reduce cognitive load by grounding abstract tasks into tangible actions (Lindgren and Johnson-Glenberg, 2013; Wilson, 2002).

In Study Two, the manipulation of virtual object sizes was explored as a secondary design element intended to enhance multimodal engagement. Although participants reported that they enjoyed the added pitch-size interaction, it was not engaged with as fully as the main pitch-colour association task. This limited use may reflect a cognitive-load constraint; introducing additional multimodal complexities can overload learners' working memory, detracting from the core learning objectives (Sweller, 2020, 2011). Users' engagement and interactions with the environment closely align with the principles of multisensory integration (as discussed in Section 2.5.2). Interactions and the feedback they produce rely on temporal synchrony to reinforce multisensory integration and congruency, which is essential for cross-modal tasks. By providing congruent and well-timed actions, these interactions can help strengthen the user's ability to form and maintain pitch-colour cross-modal associations in the VR music learning context. Consequently, careful consideration must be given to balancing interactive complexity and instructional clarity to best support cross-modal learning tasks in VR.

Iterative refinements in colour-selection interactions also led to valuable design outcomes. Study One employed a HSV colour picker; this was part of an established online test platform and, as previously highlighted, is complicated to adopt in an immersive 3D VR environment. Study Two subsequently simplified the interaction to hue selection only; however, participants requested greater flexibility. Study Three introduced a discrete thirty-five-colour palette based on Munsell colours, offering a balanced range of hues with varying saturation and brightness levels. Participants found this approach to be intuitive and satisfactory, aligning with research suggesting that VR colour pickers combining discrete and continuous methods can optimise usability and selection accuracy (Kim et al., 2022; Palmer et al., 2013). Additionally, since pitch-colour cross-modal research has highlighted connections with different colour features, such as brightness and hue, it is important to provide options for both when designing these tasks. Nonetheless, further validation is necessary to refine these interactions, determining the optimal balance between user freedom and cognitive simplicity.

Together, these insights highlight key considerations for improving usability and engagement in cross-modal VR music tasks. The following section builds on this by presenting clear design guidelines for developing future VR music-learning tools.

Design Guidelines for VR Music Tools

Some of the design features of the VR music exercises worked well, while others presented challenges for participants. Using 3D space and spatial audio helped users locate musical notes and recognise auditory sequences more easily. This aligns with previous research, showing that spatial audio and 3D spatial layouts can improve immersion, spatial awareness, and task performance in virtual environments (Potter et al., 2022; Huang et al., 2024).

The locomotion methods explored in Study Two and Study Three provided both positive and negative experiences to the participants, with it increasing engagement and immersion but also increasing simulator sickness and discomfort. With this in mind, the music exercises in the third study were conducted from a fixed position, which also supported the use of 3D audio for spatial feedback. Still, more research is needed to determine the most effective locomotion methods for VR pitch-colour association tasks. Alternative options such as walk-in-place and arm-swinging locomotion could be explored in future work, although these techniques are currently less widely adopted than traditional controller-based approaches like teleportation and free movement (Boletsis and Cedergren, 2019; Wilson et al., 2016). While such alternatives may help address comfort and motion sickness, they could also introduce new challenges, such as increased cognitive load due to the added physical effort required during learning tasks (Hoang et al., 2020).

Features like locomotion and spatial audio were not added simply to boost general immersion. Instead, they were designed to directly support pitch-colour cross-modal learning by placing sound and colour pairings into a clear spatial and embodied context. Locomotion allows users to physically move towards or interact with the sounds and colours, which helps reinforce the embodied aspect of the task. In turn, spatial audio gives precise cues that help users match the location of a note's sound source to where its colour appears, supporting multisensory integration through both spatial and temporal alignment.

As part of this PhD, a research placement at Kungliga Tekniska Högskolan (KTH) in Stockholm explored how this work could extend to other VR music applications, focusing on using an electric guitar in VR/AR environments. Although this work was still in the early stages and was limited by technological constraints, integrating cross-modal associations and real-time visual feedback from the guitar's audio output showed promising potential. Some initial research has also pointed to the benefits of augmented reality in guitar learning, with studies suggesting that real-time visual cues can support skill development (Hees, 2021). Using a cross-modal design, this immersive, feedback-driven approach could provide a fresh and engaging approach to learn the guitar, or even open up new possibilities for remote musical collaboration in creative and intuitive ways.

This section has highlighted some of the practical contributions emerging from the exploration of pitch-colour associations in VR environments. While further investigation is needed to fully realise the potential of certain design features, this work points towards several promising directions for future research and development. The following section now turns to the theoretical contributions of this thesis.

7.3.2 Theoretical Contributions

Cross-Modal Congruency in Virtual Environments

While previous research has shown that general auditory-visual associations can carry over into immersive VR (Malpica et al., 2020), it was still unclear whether the more specific link between pitch and colour would remain consistent in the immersive context of VR delivered through HMDs. The three studies presented in this thesis suggest it does; many participants consistently associated higher pitches with brighter colours and lower pitches with darker ones. This reflects patterns found in earlier research conducted outside of VR (Ward et al., 2006; Spence, 2011). In Chapter 4, results from the consistency test confirmed that these associations hold even when experienced in low-immersion using a VR HMD rather than on a traditional screen. Participant feedback in Chapters 5 and 6 also supports this, with some describing their choices as intuitive or immediate, suggesting that the immersive format did not disrupt their perceptual links. This research helps confirm that pitch-colour congruency is present and stable within VR, offering a useful connection between traditional cross-modal studies and applied VR design, and strengthening the case for using immersive environments in both perception research and multisensory learning tools.

Embodied Cognition and Multisensory Learning

This research contributes to the fields of embodied cognition and multisensory learning by showing how immersive interaction in VR can support users in forming pitch-colour cross-modal associations. Across the three studies, the findings reflect ideas from embodied cognition, which suggest that learning can be improved when abstract concepts are tied to physical actions and sensory engagement (Wilson, 2002; Lindgren and Johnson-Glenberg, 2013). Participants were asked to physically interact with objects, grabbing and placing musical orbs to match pitch with colour, which helped link sensory and motor experiences to the learning task.

The research also demonstrates how embodiment and cross-modal learning can reinforce each other. When users engage in tasks that integrate auditory, visual, and motor inputs in a congruent manner, these multisensory interactions strengthen perceptual links between modalities and support memory and conceptual understanding (Shams and Seitz, 2008). The findings from Study Three suggest that pitch-colour associations are not only perceptual phenomena but can also function as effective learning tools when embedded within embodied interactions.

At the same time, the findings highlight the importance of getting the right balance, in terms of immersion, presence, interaction design and cognitive load. This reflects similar concerns raised by Young et al. (2025) and Makransky et al. (2019), who found that higher immersion, despite boosting presence, was associated with greater cognitive load and lower learning performance. In Study Two, a secondary task was introduced where participants could adjust the size of the musical orb to match pitch. Although this was well received by participants, most did not use it during the main task. The use of the secondary task may point to cognitive overload. While the interaction was interesting, it may have added too much complexity to the task. This reflects principles from cognitive load theory, which warn that too many tasks or modes of interaction can overload working memory and reduce learning

efficiency (Sweller, 2011; Mayer, 2005).

Together, these findings show that embodied interactions in VR can support learning through cross-modal experiences, but only when designed with care. For future VR learning environments, it is essential to consider the number of interactions being requested of the user and the cognitive effort required for each task. Balancing sensory engagement with clarity and simplicity will help ensure these learning environments remain effective and accessible.

Personalisation and Cross-Modal Variability

This thesis highlights that personalised and pre-defined pitch-colour mappings produce similar learning outcomes, showing that both approaches are useful in VR educational contexts. In Study Three, participants who chose their own colour mappings performed similarly to those using a pre-defined mapping, demonstrating that allowing users to create their own associations can be just as effective as giving them a fixed set.

One of the key findings of this research is the wide variety in how participants formed their colour choices. Although general trends remained consistent, such as brighter colours linked to higher pitches, the rationale behind these choices varied widely. Some participants relied on intuition or memory, while others explicitly mentioned musical relationships or aesthetic considerations. This diversity aligns with previous findings that suggest cross-modal associations differ depending on personal experiences, musical experience, cultural background, and creative engagement (Palmer et al., 2013; Whiteford et al., 2018; O'Toole et al., 2019).

The variability observed between participants indicates the importance of flexibility in designing cross-modal VR experiences. Relying solely on either personalised or pre-defined mappings may limit the overall effectiveness for different learners. Therefore, it would be ideal to incorporate adaptable systems that provide structured mappings for beginners while allowing more experienced or creatively inclined users to customise their associations.

Such an adaptive design aligns well with educational psychology and VR design literature, emphasising that tasks should match the user's prior knowledge, cognitive style, and personal preferences (Mayer, 2005; Lindgren and Johnson-Glenberg, 2013; Marougkas et al., 2024). Overall, this research contributes valuable insights for designing cross-modal VR tools, highlighting the importance of accommodating individual differences while ensuring educational clarity and effectiveness.

Synaesthetic Continuum and Non-Synaesthetes

In Chapter 2, the idea of viewing sound-colour synaesthesia (chromaesthesia) as a continuum rather than a binary condition was introduced. Previous research has raised questions about treating synaesthesia (in certain contexts) as a binary trait, with several studies suggesting that cross-modal sensitivity exists on a spectrum (Ward et al., 2006; Campen, 2010). The findings from this thesis support the view of pitch-colour associations being on a spectrum, showing a range of strengths in how participants formed pitch-colour associations. In Chapter 4, some participants scored close to what would be expected from synaesthetes in the pitch-colour consistency test. In the more immersive studies in Chapters 5 and 6, qualitative feedback also reflected differences in the degree to which participants felt the connection

between pitch and colour. Musical background, artistic interests, and memory performance seemed to affect how stable or intuitive these associations were. Together, the findings suggest that non-synaesthetes can still form strong and consistent cross-modal links, particularly in well-designed immersive settings. By documenting graded levels of consistency rather than a strict synaesthete/non-synaesthete split, this thesis extends the continuum model into virtual environments.

This contribution works in two ways. First, it adds to the evidence that cross-modal associations exist on a spectrum and shows how immersive VR design can bring out and even strengthen these latent connections in non-synaesthetes. Second, it points to the value of designing adaptive systems that respond to where users fall along this continuum, supporting learning and engagement without requiring users to already possess strong cross-modal associations.

7.4 Can VR Enhance Cross-Modality?

Earlier in this chapter, the main themes and contributions across the three studies were discussed. Drawing from these insights, this section returns to the central research question: *"How can the cross-modal associations of pitch and colour be enhanced through virtual reality?"*

While it is challenging to assert definitively that VR enhances pitch-colour associations, the findings presented in this thesis suggest that VR has unique qualities. These qualities—such as interaction, presence, embodiment, multisensory integration, and affective design—can support and strengthen the perceptual connections between pitch and colour.

A key finding across the three studies is that pitch-colour cross-modal associations do occur within VR environments. Previous research has shown that general auditory-visual associations persist in VR, even if slightly weaker than in real-world contexts (Malpica et al., 2020), but little was known about more specific associations like pitch and colour. The results of this research go further by showing that pitch-colour associations in VR align with previously established mappings, such as higher pitches being associated with brighter colours and lower pitches with darker tones (Ward et al., 2006; Spence, 2011). This was supported both through consistency scores (see Figure 4.3) and participant feedback across the studies.

Beyond confirming that these associations can exist in VR, the research explored how immersive features might support or enhance users' understanding of the relationship between pitch and colour. The design across the studies drew on ideas from embodied cognition and multisensory integration. Participants were not simply listening to sounds or looking at colours; they were actively engaging with them through movement and interaction. Many participants reported that being able to grab the musical orbs and adjust their colour helped them stay engaged with the task and form stronger associations. These forms of interaction reflect the principles of embodied learning (Wilson, 2002), where physical actions help to support memory and understanding. Additionally, presenting auditory and visual information in temporal synchrony, a principle highlighted by Shams and Seitz (2008), appeared to support multisensory integration and helped participants focus on the task.

Design choices around the environment also played a role. Feedback from participants suggested that simpler, less visually complex environments helped them concentrate more effectively. The use of a minimal *"space"* setting in the later stud-

ies provided a focused environment that supported attention without introducing visual overload. These findings support the idea of VR as an affective medium, where emotional resonance with the virtual environment influences perception and engagement (Riva et al., 2007). As previously stated (see Section 7.1), mood and emotion can influence the kinds of associations people make between sound and colour (Palmer et al., 2013; O'Toole et al., 2019, 2021), so creating environments that support a calm and focused mindset may also help strengthen these associations.

Another design aim of this research was to explore the potential for simulating synaesthesia-like experiences for users without synaesthesia. Features such as visualised feedback to pitch-colour associations were developed to help participants *"see"* the sounds they were hearing. While not every participant experienced this similarly, many noted that this approach helped them think more deeply about their perceptual choices and made the task more engaging.

Taken together, the studies presented in this thesis suggest that VR could play a useful role in supporting pitch-colour cross-modal learning. The results show that pitch-colour associations are preserved in virtual environments and that well-designed interactions and environments can support the formation and reinforcement of these associations. While the third study focused on a music learning scenario, the findings could apply more broadly. There is clear potential to extend this research into other areas such as emotional regulation training, creative arts education, and sensory-integrated learning tools for neurodiverse users.

Although these findings are promising, they also indicate the need for further research, especially regarding longer-term use and wider user testing. These limitations are discussed next, along with suggested directions for future work.

7.5 Limitations

While each study presented in Chapters 4, 5, and 6 discusses individual limitations and future work, this section focuses on broader limitations across the thesis as a whole. These include sample size, short-term exposure, hardware and software constraints, and inconsistencies in how pitch-colour choices were presented.

One limitation across the three studies is the relatively small sample sizes: 20 participants in Study One, 38 in Study Two, and 30 in Study Three. While these numbers are within acceptable ranges for exploratory and early-stage VR research, they do limit the generalisability of the findings. For mixed methods studies, sample sizes for the quantitative components are typically recommended to fall between 30 and 50 participants to achieve sufficient statistical power (Onwuegbuzie and Collins, 2007). Usability studies in VR contexts often target around 40 participants for robust quantitative analysis (Budiu and Moran, 2021). In contrast, qualitative strands generally require fewer participants, focusing instead on depth, richness, and the point of data saturation rather than statistical generalisation. In this research, the qualitative data were used to complement and explain quantitative findings, and the sample sizes were sufficient to identify recurring themes and generate insights for future design directions.

Given the practical constraints of VR research, such as the resource-intensive nature of equipment setup and participant recruitment, these sample sizes were considered suitable for the scope and aims of the research. However, future work would

benefit from larger and more diverse participant groups to strengthen reliability and applicability.

Another key limitation was the short-term nature of participant engagement. While sessions were long enough to complete the tasks, most participants only took part in one study, with no repeated exposure over time. This limited the ability to observe long-term changes in cross-modal learning, such as the development or stability of pitch-colour associations.

Additionally, different groups of participants were used in each study, with no longitudinal tracking. As a result, it was not possible to explore how cross-modal associations evolve within individuals over time or whether prior mappings influence later performance. However, this approach also ensured that learning effects did not carry over between studies, allowing each to be assessed independently without the influence of repeated exposure or familiarity with prior tasks.

All studies were conducted using the stand-alone `Oculus Quest 2` headset and developed in `Unity` using the `XR Interaction Toolkit (XRI)`. Although this setup was robust, hardware limitations affected the participant's experience. Issues such as the weight of the headset, resolution limitations, and graphical performance—particularly in the *"forest"* environment—were commonly mentioned in feedback. These limitations affected immersion and could have impacted task focus and participant comfort.

Table D.3 outlines newer HMDs released since the `Quest 2`, many of which offer better optics, processors, and design improvements [1]. Newer headsets may offer a more comfortable and visually immersive experience, which could help reduce some of the issues encountered.

On the software side, the project used version 2.x of `Unity's XR Interaction Toolkit (XRI)`, which relied on older input and interaction systems. Although version 3.x offered several improvements, a full upgrade was not possible due to time and integration constraints. This means that some newer interaction features were unavailable for this project, and future iterations could benefit from adopting these updated tools.

Another limitation is the variation in how pitch-colour choices were presented across the studies. Study One used a full colour picker with thirteen notes, while Study Two reduced both the note range to eight and limited colour selection to basic hue categories. Study Three used a thirty-five-colour palette and the same eight notes used in Study Two (C Major Scale). These changes reflected attempts to improve usability and task clarity, but also complicate cross-study comparison. Differences in results may partly reflect changes in design rather than just user perception.

Despite these limitations, the iterative design process helped refine the interaction and presentation of pitch-colour mapping tasks. Future work could aim for consistency across studies to better compare outcomes while refining and improving the user experience.

[1] VRcompare, `https://vr-compare.com`, accessed 25 May 2025.

7.6 Future Work

Looking ahead, there are several areas where this research could be expanded. One key direction is conducting longitudinal studies that track how cross-modal associations develop over multiple sessions. With the foundations established in this thesis, future studies could explore whether pitch-colour associations remain stable or shift over time and how engagement evolves with repeated exposure.

Another promising avenue is exploring AI-driven personalisation. Future systems could adapt pitch-colour mappings based on a user's response times, interaction patterns, personality traits, and emotional feedback. Building on earlier work from O'Toole et al. (2019) and O'Toole et al. (2021) (P1), this approach could help tailor cross-modal tasks to each user's perceptual and cognitive profile. For instance, users who show difficulty with certain mappings could be supported with alternative prompts or visual feedback, while users with high consistency might be encouraged to deepen their associations through more complex interactions. Over time, such a system could refine its suggestions using patterns found across similar users, offering a flexible and inclusive learning tool.

Expanding the research beyond pitch and colour could also yield valuable insights. Other auditory and visual features—such as pitch-size (Gallace and Spence, 2006), pitch-shape (Marks, 1987), loudness-brightness (Marks, 1987), and timbre-shape (Gurman et al., 2021)—could be tested to see which associations users naturally gravitate toward in VR. Combining multiple cross-modal pairings might enrich learning environments, offering layered cues that support memory and understanding.

Beyond education, this research could be applied to various creative and therapeutic contexts. For example, cross-modal VR tools could support music-based therapy by enabling non-verbal emotional expression through colour and sound. Applications for neurodivergent users, including those with autism or Attention-Deficit/Hyperactivity Disorder (ADHD), could also be explored, where structured and personalised VR experiences help with focus, engagement, and sensory regulation.

Creative use cases, such as painting with sound or collaborative music composition in immersive environments, represent further directions for this research. These ideas build on this thesis's immersive, interactive foundations and suggest how pitch-colour mapping can support a broader range of user experiences.

Together, these future directions provide a roadmap for extending the impact of this research, from refining educational VR tools to developing new applications that support creativity, emotion, and accessibility.

7.7 Conclusion

This thesis aimed to explore the question: *"How can the cross-modal associations of pitch and colour be enhanced through virtual reality?"*

While much research has focused on VR as an effective learning platform, very little has examined cross-modality—especially the connection between pitch and colour—within virtual environments. Chapters 4, 5, and 6 presented three studies that sought to answer this central research question. Study One explored whether VR HMDs were a suitable platform for pitch-colour associations. Study Two exam-

ined how immersive features of VR—such as environment, audio-visual, and interaction design—could support perceptual connections. Finally, Study Three tested a learning environment designed to use pitch-colour associations in a music theory context.

Across these three studies, it was found that pitch-colour cross-modal associations do occur in virtual environments, in ways consistent with real-world settings. To strengthen these associations, certain design features—such as direct interaction, emotional tone, and adaptive feedback—proved particularly effective. These findings also highlight that both personalisation and carefully designed pre-defined mappings can be effective, and that interaction and affective engagement help support a deeper understanding of cross-modal relationships.

This thesis contributes to the understanding of cross-modal perception while also informing the design of VR-based learning environments. Although the use case in Study Three focused on music theory, the findings suggest wider potential applications. Music therapy, neurodivergent learning, and creative arts education could benefit from thoughtful cross-modal design in immersive contexts.

The insights gained here offer a glimpse into how future educational and therapeutic systems might draw on the full range of human perception, opening new ways of learning, expressing, and connecting across the senses.

Bibliography

Adeli, M., Rouat, J. and Molotchnikoff, S. (2014), 'Audiovisual correspondence between musical timbre and visual shapes', *Frontiers in Human Neuroscience* **8**. https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2014.00352/full

Ahn, J. and Clegg, T. (2018), Human-Computer Interaction and Education: Designing for Technology-Enhanced Learning Experiences, *in* K. L. Norman and J. Kirakowski, eds, 'The Wiley Handbook of Human Computer Interaction', 1 edn, Wiley, pp. 821–830. https://onlinelibrary.wiley.com/doi/10.1002/9781118976005.ch38

Alase, A. (2017), 'The Interpretative Phenomenological Analysis (IPA): A Guide to a Good Qualitative Research Approach', *International Journal of Education and Literacy Studies* **5**(2), 9. https://eric.ed.gov/?id=EJ1149107

Alwashmi, K., Meyer, G., Rowe, F. and Ward, R. (2024), 'Enhancing learning outcomes through multisensory integration: A fMRI study of audio-visual training in virtual reality', *NeuroImage* **285**, 120483. https://www.sciencedirect.com/science/article/pii/S105381192300633X

Asano, M. and Yokosawa, K. (2011), 'Synesthetic colors are elicited by sound quality in Japanese synesthetes', *Consciousness and Cognition* **20**(4), 1816–1823. http://linkinghub.elsevier.com/retrieve/pii/S1053810011001450

Asher, J., Aitken, M., Farooqi, N., Kurmani, S. and Baroncohen, S. (2006), 'Diagnosing and Phenotyping Visual Synaesthesia: a Preliminary Evaluation of the Revised Test of Genuineness (TOG-R)', *Cortex* **42**(2), 137–146. https://linkinghub.elsevier.com/retrieve/pii/S001094520870337X

Balcerak Jackson, M. and Balcerak Jackson, B. (2024), 'Immersive Experience and Virtual Reality', *Philosophy & Technology* **37**(1), 19. https://doi.org/10.1007/s13347-024-00707-1

Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J. and Bolton, P. (1996), 'Synaesthesia: Prevalence and Familiality', *Perception* **25**(9), 1073–1079. https://doi.org/10.1068/p251073

Baron-Cohen, S., Harrison, J., Goldstein, L. H. and Wyke, M. (1993), 'Coloured Speech Perception: Is Synaesthesia what Happens when Modularity Breaks Down?', *Perception* **22**(4), 419–426. https://doi.org/10.1068/p220419

Baron-Cohen, S., Wyke, M. A. and Binnie, C. (1987), 'Hearing Words and Seeing Colours: An Experimental Investigation of a Case of Synaesthesia', *Perception* **16**(6), 761–767.
https://doi.org/10.1068/p160761

Belloli, S., Virk, V., Etzi, R., Gallace, A., Bordegoni, M., Carulli, M. and Porro, S. (2020), 'The Kandinsky Experience: A Multisensory Augmented Reality Application for Cultural Heritage', *Computer-Aided Design and Applications* **18**(4), 799–814.
https://hdl.handle.net/11311/1168706

Ben-Artzi, E. and Marks, L. E. (1995), 'Visual-auditory interaction in speeded classification: Role of stimulus difference', *Perception & Psychophysics* **57**(8), 1151–1162. Place: US.
https://doi.org/10.3758/BF03208371

Bermudez i Badia, S., Quintero, L. V., Cameirao, M. S., Chirico, A., Triberti, S., Cipresso, P. and Gaggioli, A. (2019), 'Toward Emotionally Adaptive Virtual Reality for Mental Health Applications', *IEEE Journal of Biomedical and Health Informatics* **23**(5), 1877–1887.
https://ieeexplore.ieee.org/document/8516285/

Bernstein, I. H., Eason, T. R. and Schurman, D. L. (1971), 'Hue-tone sensory interaction: A negative result', *Perceptual and Motor Skills* **33**(3, Pt. 2), 1327–1330. Place: US.
https://doi.org/10.2466/pms.1971.33.3f.1327

Bernstein, I. H. and Edelstein, B. A. (1971), 'Effects of some variations in auditory input upon visual choice reaction time', *Journal of Experimental Psychology* **87**(2), 241–247. Place: US.
https://doi.org/10.1037/h0030524

Bhatti, A., Behinaein, B., Rodenburg, D., Hungler, P. and Etemad, A. (2021), Attentive Cross-modal Connections for Deep Multimodal Wearable-based Emotion Recognition, *in* '2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)', pp. 01–05.
https://ieeexplore.ieee.org/document/9666360/?arnumber=9666360

Billger, M., Heldal, I., Stahre, B. and Renstrom, K. (2004), Perception of color and space in virtual reality: a comparison between a real room and virtual reality models, *in* B. E. Rogowitz and T. N. Pappas, eds, 'The International Society for Optical Engineering', San Jose, CA, p. 90.
http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.526986

Boletsis, C. and Cedergren, J. E. (2019), 'VR Locomotion in the New Era of Virtual Reality: An Empirical Comparison of Prevalent Techniques', *Advances in Human-Computer Interaction* **2019**, e7420781.
https://www.hindawi.com/journals/ahci/2019/7420781/

Bordegoni, M., Carulli, M. and Spadoni, E. (2023), Applications of Multisensory Experiences in eXtended Reality, *in* M. Bordegoni, M. Carulli and E. Spadoni, eds, 'Prototyping User eXperience in eXtended Reality', SpringerBriefs in Applied Sciences and Technology, Springer Nature Switzerland, Cham, pp. 93–112.
https://doi.org/10.1007/978-3-031-39683-0_7

Bowman, D. A., Gabbard, J. L. and Hix, D. (2002), 'A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods', *Presence: Teleoperators and Virtual Environments* **11**(4), 404–424.
https://doi.org/10.1162/105474602760204309

Bowman, D. A., Kruijff, E., LaViola, J. J. and Poupyrev, I. (2001), 'An Introduction to 3-D User Interface Design', *Presence: Teleoperators and Virtual Environments* **10**(1), 96–108.
https://direct.mit.edu/pvar/article/10/1/96-108/18291

Bowman, D. A. and McMahan, R. P. (2007), 'Virtual Reality: How Much Immersion Is Enough?', *Computer* **40**(7), 36–43. Conference Name: Computer.
https://ieeexplore.ieee.org/abstract/document/4287241/

Brang, D. and Ramachandran, V. S. (2011), 'Survival of the Synesthesia Gene: Why Do People Hear Colors and Taste Words?', *PLoS Biology* **9**(11), e1001205.
http://dx.plos.org/10.1371/journal.pbio.1001205

Braun, V. and Clarke, V. (2019), 'Reflecting on reflexive thematic analysis', *Qualitative Research in Sport, Exercise and Health* **11**(4), 589–597. _eprint: https://doi.org/10.1080/2159676X.2019.1628806.
https://doi.org/10.1080/2159676X.2019.1628806

Bresin, R. (2005), 'What is the color of that music performance?', *Proceedings of the International Computer Music Conference–ICMC* pp. 367–370.
https://www.researchgate.net/profile/Roberto-Bresin/publication/248421913_What_is_the_color_of_that_music_performance/links/004635209648ddd7e5000000/What-is-the-color-of-that-music-performance.pdf

Brinkman Willem-Paul, Hoekstra Allart R.D. and Van Egmond René (2015), The Effect Of 3D Audio And Other Audio Techniques On Virtual Reality Experience, *in* 'Studies in Health Technology and Informatics', IOS Press.
https://www.medra.org/servlet/aliasResolver?alias=iospressISBN&isbn=978-1-61499-594-4&spage=44&doi=10.3233/978-1-61499-595-1-44

Broderick, J. (2021), Effect of Spatial Audio and Movement Control on User Performance, Awareness, and Preference, PhD thesis, University of Galway, Galway, Ireland.
https://doi.org/10.13025/17547

Broderick, J., Duggan, J. and Redfern, S. (2018), The Importance of Spatial Audio in Modern Games and Virtual Environments, *in* '2018 IEEE Games, Entertainment, Media Conference (GEM)', pp. 1–9.
https://ieeexplore.ieee.org/document/8516445

Brogaard, B. (2020), Colour synaesthesia and its philosophical implications, *in* 'The Routledge Handbook of Philosophy of Colour', Routledge.

Brooke, J. (1996), SUS: A 'Quick and Dirty' Usability Scale, *in* P. W. Jordan, B. Thomas, I. L. McClelland and B. Weerdmeester, eds, 'Usability Evaluation In Industry', 0 edn, CRC Press, pp. 207–212.
https://www.taylorfrancis.com/books/9781498710411/chapters/10.1201/9781498710411-35

Brooks, F. (1999), 'What's real about virtual reality?', *IEEE Computer Graphics and Applications* **19**(6), 16–27. Conference Name: IEEE Computer Graphics and Applications.
https://ieeexplore.ieee.org/document/799723

Brooks, J. and King, N. (2012), Qualitative psychology in the real world: The utility of template analysis, *in* 'British Psychological Society Annual Conference', London, UK.
https://eprints.hud.ac.uk/id/eprint/13656/1/Brooks__King_QMiP_2012_Final.pdf

Brooks, J., McCluskey, S., Turley, E. and King, N. (2015), 'The Utility of Template Analysis in Qualitative Psychology Research', *Qualitative Research in Psychology* **12**(2), 202–222.
https://doi.org/10.1080/14780887.2014.955224

Brown, P., Spronck, P. and Powell, W. (2022), 'The simulator sickness questionnaire, and the erroneous zero baseline assumption', *Frontiers in Virtual Reality* **3**.
https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2022.945800/full

Bu, L., Cui, B., Pan, W., Chen, H., Xia, S. and Li, H. (2025), 'Towards Inclusive Autonomous Vehicles: Multimodal Interaction Design for Enhancing Accessibility and Reducing Cognitive Load in Drivers and Passengers with Severe Visual Impairment'.
https://papers.ssrn.com/abstract=5085389

Budiu, R. and Moran, K. (2021), 'How Many Participants for Quantitative Usability Studies: A Summary of Sample-Size Recommendations'.
https://www.nngroup.com/articles/summary-quant-sample-sizes/

Calvert, G., Spence, C. and Stein, B. E., eds (2004), *The handbook of multisensory processes*, MIT Press, Cambridge, Mass.

Campen, C. v. (2010), *The hidden sense: synesthesia in art and science*, MIT Press, Cambridge, Mass. OCLC: 906956300.

Campen, C. v. and Froger, C. (2003), 'Personal Profiles of Color Synesthesia: Developing a Testing Method for Artists and Scientists', *Leonardo* **36**(4), 291–294.
https://direct.mit.edu/leon/article/36/4/291-294/44443

Capece, N., Erra, U., Malandrino, D., North, M. M. and Gruosso, M. (2023), 'Evaluation of Virtual Reality Interaction Techniques: the case of 3D Graph'. arXiv:2302.05660 [cs].
http://arxiv.org/abs/2302.05660

Carmichael, D., Down, M., Shillcock, R., Eagleman, D. and Simner, J. (2015), 'Validating a standardised test battery for synesthesia: Does the Synesthesia Battery reliably detect synesthesia?', *Consciousness and Cognition* **33**, 375–385.
https://linkinghub.elsevier.com/retrieve/pii/S1053810015000276

Chalmers, A., Debattista, K. and Ramic-Brkic, B. (2009), 'Towards high-fidelity multi-sensory virtual environments', *The Visual Computer* **25**(12), 1101–1108.
https://doi.org/10.1007/s00371-009-0389-2

Chamorro-Premuzic, T. and Furnham, A. (2007), 'Personality and music: Can traits explain how people use music in everyday life?', *British Journal of Psychology* **98**(2), 175–185. _eprint: https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1348/000712606X111177.
https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1348/000712606X111177

Chen, H.-H. and Tsao, H.-Y. (2023), A Framework For Presence Maintenance: Evoking Synesthesia and Empathy in Cinematic Virtual Reality, *in* '2023 IEEE Gaming, Entertainment, and Media Conference (GEM)', pp. 1–4.
https://ieeexplore.ieee.org/abstract/document/10389829

Chen, J., Dai, J., Zhu, K. and Xu, L. (2022), 'Effects of extended reality on language learning: A meta-analysis', *Frontiers in Psychology* **13**.
https://doi.org/10.3389/fpsyg.2022.1016519

Chen, Y.-C. and Spence, C. (2010), 'When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures', *Cognition* **114**(3), 389–404.
https://www.sciencedirect.com/science/article/pii/S0010027709002649

Cherni, H., Métayer, N. and Souliman, N. (2020), 'Literature review of locomotion techniques in virtual reality. | International Journal of Virtual Reality | EBSCOhost', *International Journal of Virtual Reality* **20**(1), 1–20.
https://openurl.ebsco.com/EPDB%3Agcd%3A11%3A15211003/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A142484303&crl=c

Chirico, A., Ferrise, F., Cordella, L. and Gaggioli, A. (2018), 'Designing Awe in Virtual Reality: An Experimental Study', *Frontiers in Psychology* **8**.
https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.02351/full

Clark, H. H. and Brownell, H. H. (1976), 'Position, direction, and their perceptual integrality', *Perception & Psychophysics* **19**(4), 328–334.
https://doi.org/10.3758/BF03204238

Corbin, J. and Strauss, A. (2014), *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, SAGE Publications. Google-Books-ID: MaKWBQAAQBAJ.

Creswell, J. W. (2022), *A concise introduction to mixed methods research*, second edition edn, SAGE, Los Angeles London New Delhi Singapore Washington DC Melbourne.

Cummings, J. J. and Bailenson, J. N. (2016), 'How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence', *Media Psychology* **19**(2), 272–309.
https://doi.org/10.1080/15213269.2015.1015740

Curwen, C. (2018), 'Music-colour synaesthesia: Concept, context and qualia', *Consciousness and Cognition* **61**, 94–106.
https://linkinghub.elsevier.com/retrieve/pii/S1053810017305883

Curwen, C., Timmers, R. and Schiavio, A. (2023), 'Action, emotion, and music-colour synaesthesia: an examination of sensorimotor and emotional responses in synaesthetes and non-synaesthetes', *Psychological Research* .
https://link.springer.com/10.1007/s00426-023-01856-2

Cytowic, R. E. (1989), *Synesthesia: a union of the senses*, Springer-Verlag, New York. OCLC: 1105389247.

Cytowic, R. E. (2003), *The man who tasted shapes*, MIT Press, Cambridge, Mass. OCLC: 937348414.

Cytowic, R. E. (2012), *Synesthesia: A Union of the Senses*, Springer Science & Business Media. Google-Books-ID: yoy1BwAAQBAJ.

Day, S. A. (2022), 'Types-of-Syn'.
http://www.daysyn.com/Types-of-Syn.html

Dede, C. (2009), 'Immersive Interfaces for Engagement and Learning', *Science (American Association for the Advancement of Science)* **323**(5910), 66–69. Place: United States.
https://www.science.org/doi/10.1126/science.1167311

Deroy, O. and Spence, C. (2013), 'Why we are not all synesthetes (not even weakly so)', *Psychonomic Bulletin & Review* **20**(4), 643–664.
http://link.springer.com/10.3758/s13423-013-0387-2

Deroy, O. and Spence, C. (2016), 'Crossmodal Correspondences: Four Challenges', *Multisensory Research* **29**(1-3), 29–48.
https://doi.org/10.1163/22134808-00002488

Díaz-Barrancas, F., Cwierz, H., Pardo, P. J., Pérez, A. L. and Suero, M. I. (2020), 'Spectral Color Management in Virtual Reality Scenes', *Sensors* **20**(19), 5658.
https://www.mdpi.com/1424-8220/20/19/5658

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D. and Sarma, A. K. (2007), 'A standardized test battery for the study of synesthesia', *Journal of Neuroscience Methods* **159**(1), 139–145.
https://linkinghub.elsevier.com/retrieve/pii/S0165027006003153

Evans, K. K. and Treisman, A. (2010), 'Natural cross-modal mappings between visual and auditory features', *Journal of Vision* **10**(1), 6.
https://doi.org/10.1167/10.1.6

Farina, F. R., Mitchell, K. J. and Roche, R. A. P. (2017), 'Synaesthesia lost and found: two cases of person- and music-colour synaesthesia', *European Journal of Neuroscience* **45**(3), 472–477.
http://doi.wiley.com/10.1111/ejn.13492

Feilzer, M. (2010), 'Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm', *Journal of Mixed Methods Research* **4**(1), 6–16.
https://doi.org/10.1177/1558689809349691

Felton, W. M. and Jackson, R. E. (2022), 'Presence: A Review', *International Journal of Human–Computer Interaction* **38**(1), 1–18. _eprint: https://doi.org/10.1080/10447318.2021.1921368.
https://doi.org/10.1080/10447318.2021.1921368

Frommel, J., Sonntag, S. and Weber, M. (2017), Effects of controller-based locomotion on player experience in a virtual reality exploration game, *in* 'Proceedings of the 12th International Conference on the Foundations of Digital Games', FDG '17, Association for Computing Machinery, New York, NY, USA, pp. 1–6.
https://doi.org/10.1145/3102071.3102082

Gallace, A. and Spence, C. (2006), 'Multisensory synesthetic interactions in the speeded classification of visual size', *Perception & Psychophysics* **68**(7), 1191–1203.
https://doi.org/10.3758/BF03193720

Gallagher, S. and Lindgren, R. (2015), 'Enactive Metaphors: Learning Through Full-Body Engagement', *Educational Psychology Review* **27**(3), 391–404.
https://doi.org/10.1007/s10648-015-9327-1

Gil Rodríguez, R., Bayer, F., Toscani, M., Guarnera, D., Guarnera, G. C. and Gegenfurtner, K. R. (2022), 'Colour Calibration of a Head Mounted Display for Colour Vision Research Using Virtual Reality', *SN Computer Science* **3**(1), 22.
https://link.springer.com/10.1007/s42979-021-00855-7

Gkoumas, C. and Shimi, A. (2022), "The more you explore, the less you remember": unraveling the effects of scene clutter on learning and memory for targets, *in* '2022 Symposium on Eye Tracking Research and Applications', ETRA '22, Association for Computing Machinery, New York, NY, USA, pp. 1–7.
https://dl.acm.org/doi/10.1145/3517031.3529623

Glenn, C. P. and Coxon, M. (2024), 'Individual differences in processing multisensory information predict presence in different virtual reality environments', *Virtual Reality* **29**(1), 9.
https://doi.org/10.1007/s10055-024-01086-w

Goller, A. I., Otten, L. J. and Ward, J. (2009), 'Seeing Sounds and Hearing Colors: An Event-related Potential Study of Auditory–Visual Synesthesia', *Journal of Cognitive Neuroscience* **21**(10), 1869–1881.
http://www.mitpressjournals.org/doi/10.1162/jocn.2009.21134

Grabowski, A. and Jach, K. (2021), 'The use of virtual reality in the training of professionals: with the example of firefighters', *Computer Animation and Virtual Worlds* **32**(2), e1981. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cav.1981.
https://onlinelibrary.wiley.com/doi/abs/10.1002/cav.1981

Gradl, S., Wirth, M., Mächtlinger, N., Poguntke, R., Wonner, A., Rohleder, N. and Eskofier, B. M. (2019), The Stroop Room: A Virtual Reality-Enhanced Stroop Test, *in* '25th ACM Symposium on Virtual Reality Software and Technology', ACM, Parramatta NSW Australia, pp. 1–12.
https://dl.acm.org/doi/10.1145/3359996.3364247

Greated, M. (2011), 'The nature of sound and vision in relation to colour', *Optics & Laser Technology* **43**(2), 337–347.
https://linkinghub.elsevier.com/retrieve/pii/S0030399209001261

Greene, J. and D'Oliveira, M. (2011), *Learning to use statistical tests in psychology*, 3., ed. reprinted edn, Open University Press, Maidenhead.

Grimshaw-Aagaard, M. and Walther-Hansen, M. (2024), 'Less-is-more: auditory strategies for reduced reality', *Personal and Ubiquitous Computing* .
https://doi.org/10.1007/s00779-024-01808-6

Gröhn, M., Lokki, T. and Takala, T. (2005), 'Comparison of auditory, visual, and audiovisual navigation in a 3D space', *ACM Transactions on Applied Perception* **2**(4), 564–570.
https://doi.org/10.1145/1101530.1101558

Guo, Q. and Jiang, T. (2023), Using Crossmodal Correspondence Between Colors and Music to Enhance Online Art Exhibition Visitors' Experience, *in* I. Sserwanga, A. Goulding, H. Moulaison-Sandy, J. T. Du, A. L. Soares, V. Hessami and R. D. Frank, eds, 'Information for a Better World: Normality, Virtuality, Physicality, Inclusivity', Lecture Notes in Computer Science, Springer Nature Switzerland, Cham, pp. 144–159.
https://doi.org/10.1007/978-3-031-28035-1_12

Gurman, D., McCormick, C. R. and Klein, R. M. (2021), 'Crossmodal Correspondence Between Auditory Timbre and Visual Shape', *Multisensory Research* **35**(3), 221–241.
https://doi.org/10.1163/22134808-bja10067

Gómez Sirvent, J. L., Fernández-Sotos, D., Sánchez-Reolid, R., de la Rosa López, F., Fernández-Sotos, A. and Fernández-Caballero, A. (2023), 'Pre-occupancy evaluation of a virtual music school classroom: Influence of color and type of lighting on music performers', *Building and Environment* **246**, 110989.
https://www.sciencedirect.com/science/article/pii/S0360132323010168

Hall, J. N. (2013), 'Pragmatism, Evidence, and Mixed Methods Evaluation', *New Directions for Evaluation* **2013**(138), 15–26. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ev.20054.
https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.20054

Hart, S. G. and Staveland, L. E. (1988), Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research, *in* P. A. Hancock and N. Meshkati, eds, 'Advances in Psychology', Vol. 52 of *Human Mental Workload*, North-Holland, pp. 139–183.
https://www.sciencedirect.com/science/article/pii/S0166411508623869

Haverkamp, M. (2022), Scope and limits of synaesthetic arts, *in* 'Proceedings VII International Congress "Synaesthesia: Science and Art" * The digital / physical challenge *', Granada & Alcalà la Real / Spain.
https://www.researchgate.net/profile/Michael-Haverkamp/publication/365790044_Scope_and_limits_of_synaesthetic_arts/links/644c17b44af788735245a7db/Scope-and-limits-of-synaesthetic-arts.pdf

Hees, J. (2021), Augmented reality feedback visualization for guitars in music education, bachelorThesis, University of Stuttgart. Accepted: 2022-09-27T13:50:06Z.
http://elib.uni-stuttgart.de/handle/11682/12424

Hevner, K. (1936), 'Experimental Studies of the Elements of Expression in Music', *The American Journal of Psychology* **48**(2), 246.
https://doi.org/10.2307/1415746

Hill, L. C. (2006), Synesthetic Music Experience Communicator, PhD thesis, Digital Repository, Iowa State University.
https://lib.dr.iastate.edu/rtd/3077/

Hirci, N., Lesar, Z., Marolt, M. and Pesek, M. (2023), Evaluating Virtual Reality as Immersive Medium for Enhancing Music Skills, *in* L. T. De Paolis, P. Arpaia and M. Sacco, eds, 'Extended Reality', Lecture Notes in Computer Science, Springer Nature Switzerland, Cham, pp. 69–78.
https://doi.org/10.1007/978-3-031-43404-4_5

Hirway, A., Qiao, Y. and Murray, N. (2022), Spatial audio in 360 Degree videos: does it influence visual attention?, *in* 'Proceedings of the 13th ACM Multimedia Systems Conference', MMSys '22, Association for Computing Machinery, New York, NY, USA, pp. 39–51.
https://doi.org/10.1145/3524273.3528179

Hoang, I., Ranchet, M., Derollepot, R., Moreau, F. and Paire-Ficout, L. (2020), 'Measuring the Cognitive Workload During Dual-Task Walking in Young Adults: A Combination of Neurophysiological and Subjective Measures', *Frontiers in Human Neuroscience* **14**.

https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2020.592532/full

Hosek, A. (2023), 'Music in VR: Exploring Creative Expression and Virtually "Tangible" Learning - ProQuest'.
https://www.proquest.com/docview/2819952937?fromopenview=true&pq-origsite=gscholar&sourcetype=Dissertations%20&%20Theses

Huang, G., Chen, C., Tang, Y., Zhang, H., Liu, R. and Zhou, L. (2024), 'A study on the effect of different channel cues on learning in immersive 360° videos', *Frontiers in Psychology* **15**.
https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1335022/full

Interaction Techniques in VR (n.d.).
https://www.futurelearn.com/info/blog

Jacko, J. A. and Sears, A., eds (2003), *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*, Human factors and ergonomics, Erlbaum Associates, Mahwah, NJ.

Janata, P., , Stefan T., T., and Rakowski, S. K. (2007), 'Characterisation of music-evoked autobiographical memories', *Memory* **15**(8), 845–860. _eprint: https://doi.org/10.1080/09658210701734593.
https://doi.org/10.1080/09658210701734593

Janik McErlean, A. B. and Banissy, M. J. (2017), 'Color Processing in Synesthesia: What Synesthesia Can and Cannot Tell Us About Mechanisms of Color Processing', *Topics in Cognitive Science* **9**(1), 215–227.
http://doi.wiley.com/10.1111/tops.12237

Jerald, J. (2015), *The VR book: Human-centered design for virtual reality*, Morgan & Claypool.

Jerald, J. (2018), 'Human-Centered VR Design: Five Essentials Every Engineer Needs to Know', *IEEE Computer Graphics and Applications* **38**(2), 15–21.
https://ieeexplore.ieee.org/document/8336839

Johnson, D., Allison, C. and Baron-Cohen, S. (2013), *The prevalence of synesthesia*, Oxford University Press.
http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199603329.001.0001/oxfordhb-9780199603329-e-001

Kailas, G. and Tiwari, N. (2021), Design for Immersive Experience: Role of Spatial Audio in Extended Reality Applications, *in* A. Chakrabarti, R. Poovaiah, P. Bokil and V. Kant, eds, 'Design for Tomorrow—Volume 2', Smart Innovation, Systems and Technologies, Springer, Singapore, pp. 853–863.
https://doi.org/10.1007/978-981-16-0119-4_69

Kalloniatis, M. and Luu, C. (1995), The Perception of Color, *in* H. Kolb, E. Fernandez, B. Jones and R. Nelson, eds, 'Webvision: The Organization of the Retina and Visual System', University of Utah Health Sciences Center, Salt Lake City

(UT).
http://www.ncbi.nlm.nih.gov/books/NBK11538/

Kandinsky, W. (2022), *Concerning the spiritual in art*, Cedar Lake Classics, Erscheinungsort nicht ermittelbar.

Kennedy, R. S., Lane, N. E., Berbaum, K. S. and Lilienthal, M. G. (1993), 'Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness', *The International Journal of Aviation Psychology* **3**(3), 203–220.
https://doi.org/10.1207/s15327108ijap0303_3

Kilteni, K., Groten, R. and Slater, M. (2012), 'The Sense of Embodiment in Virtual Reality', *Presence: Teleoperators and Virtual Environments* **21**(4), 373–387.
https://direct.mit.edu/pvar/article/21/4/373-387/18838

Kim, H. K., Park, J., Choi, Y. and Choe, M. (2018), 'Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment', *Applied Ergonomics* **69**, 66–73.
https://www.sciencedirect.com/science/article/pii/S000368701730282X

Kim, J., Hwang, J.-I. and Lee, J. (2022), 'VR Color Picker: Three-Dimensional Color Selection Interfaces', *IEEE Access* **10**, 1–1.
https://ieeexplore.ieee.org/document/9800763

Kim, J., Lee, J., Kim, Y. J. and Lee, J. (2023), Influence of Cross-Modal Correspondence between Auditory and Visual Stimuli on Vection Perception in Virtual Reality, *in* '2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)', pp. 604–608.
https://ieeexplore.ieee.org/document/10322233

Kizony, R., Levin, M. F., Hughey, L., Perez, C. and Fung, J. (2010), 'Cognitive load and dual-task performance during locomotion poststroke: a feasibility study using a functional virtual environment', *Physical Therapy* **90**(2), 252–260.
https://doi.org/10.2522/ptj.20090061

Kojić, T., Vergari, M., Knuth, S., Warsinke, M., Möller, S. and Voigt-Antons, J.-N. (2024), Influence of Gameplay Duration, Hand Tracking, and Controller Based Control Methods on UX in VR, *in* 'Proceedings of the 16th International Workshop on Immersive Mixed and Virtual Environment Systems', pp. 22–28. arXiv:2404.03337 [cs].
http://arxiv.org/abs/2404.03337

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T. and White-Schwoch, T. (2014), 'Auditory learning through active engagement with sound: biological impact of community music lessons in at-risk children', *Frontiers in Neuroscience* **8**, 1–12.
http://journal.frontiersin.org/article/10.3389/fnins.2014.00351/abstract

Kuhail, M. A., ElSayary, A., Farooq, S. and Alghamdi, A. (2022), 'Exploring Immersive Learning Experiences: A Survey', *Informatics* **9**(4), 75. Number: 4.
https://www.mdpi.com/2227-9709/9/4/75

Kuo, Y.-T. and Chuang, M.-C. (2013), 'A proposal of a color music notation system on a single melody for music beginners', *International Journal of Music Education* **31**(4), 394–412.
http://journals.sagepub.com/doi/10.1177/0255761413489082

Lacey, S., Martinez, M., McCormick, K. and Sathian, K. (2016), 'Synesthesia strengthens sound-symbolic cross-modal correspondences', *European Journal of Neuroscience* **44**(9), 2716–2721.
https://onlinelibrary.wiley.com/doi/10.1111/ejn.13381

Larsson, P., Västfjäll, D. and Kleiner, M. (2002), Auditory-Visual Interaction in Real and Virtual Rooms, *in* 'Proceedings of the Forum Acusticum', Vol. 23, Sevilla, Spain, p. 6.
https://documentacion.sea-acustica.es/storage/publicaciones/Sevilla02_psy05004.pdf

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H. and Wallace, M. T. (2004), 'Semantic congruence is a critical factor in multisensory behavioral performance', *Experimental Brain Research* **158**(4), 405–414.
https://doi.org/10.1007/s00221-004-1913-2

LaViola, J. J., Kruijff, E., McMahan, R. P., Bowman, D. and Poupyrev, I. P. (2017), *3D User Interfaces: Theory and Practice*, Addison-Wesley Professional.

Lee, C. H., Lockton, D., Stevens, J., Wang, S. J. and Ahn, S. (2019), Synaesthetic-Translation Tool: Synaesthesia as an Interactive Material for Ideation, *in* 'Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems', ACM, Glasgow Scotland Uk, pp. 1–6.
https://dl.acm.org/doi/10.1145/3290607.3312849

Lee, K. M. (2004), 'Presence, Explicated', *Communication Theory* **14**(1), 27–50.
https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lee, M., Kim, E., Choe, J., Choi, S., Ha, S. and Kim, G. (2022), 'Psychological Effects of Green Experiences in a Virtual Environment: A Systematic Review', *Forests* **13**(10), 1625. Number: 10.
https://www.mdpi.com/1999-4907/13/10/1625

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T. and Van Merriënboer, J. J. G. (2013), 'Development of an instrument for measuring different types of cognitive load', *Behavior Research Methods* **45**(4), 1058–1072.
https://doi.org/10.3758/s13428-013-0334-1

Lewis, J. (2020), 'How children listen: multimodality and its implications for K-12 music education and music teacher education', *Music Education Research* **22**(4), 373–387.
https://www.tandfonline.com/doi/full/10.1080/14613808.2020.1781804

Lewis, J. R. (2018), 'The System Usability Scale: Past, Present, and Future', *International Journal of Human–Computer Interaction* **34**(7), 577–590.
https://www.tandfonline.com/doi/full/10.1080/10447318.2018.1455307

Lindgren, R. and Johnson-Glenberg, M. (2013), 'Emboldened by Embodiment: Six Precepts for Research on Embodied Learning and Mixed Reality', *Educational Researcher* **42**(8), 445–452.
https://doi.org/10.3102/0013189X13511661

Lombard, M. and Jones, M. T. (2015), Defining Presence, *in* M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn and R. J. Schaevitz, eds, 'Immersed in Media: Telepresence Theory, Measurement & Technology', Springer International Publishing, Cham, pp. 13–34.
https://doi.org/10.1007/978-3-319-10190-3_2

Lorenzo, G., Gilabert Cerdá, A., Lorenzo-Lledó, A. and Lledó, A. (2022), 'The application of augmented reality in the learning of autistic students: a systematic and thematic review in 1996–2020', *Journal of Enabling Technologies* **16**(2), 75–90.
https://www.emerald.com/insight/content/doi/10.1108/JET-12-2021-0068/full/html

Lu, Y., Wang, X., Gong, J. and Liang, Y. (2022), 'ChordAR: An Educational AR Game Design for Children's Music Theory Learning', *Wireless Communications and Mobile Computing* **2022**, e5268586.
https://www.hindawi.com/journals/wcmc/2022/5268586/

Luong, T., Lecuyer, A., Martin, N. and Argelaguet, F. (2021), 'A Survey on Affective and Cognitive VR', *IEEE Transactions on Visualization and Computer Graphics* pp. 1–1.
https://ieeexplore.ieee.org/document/9531381/

Macpherson, F. (2007), Synaesthesia, Functionalism and Phenomenology, *in* M. Marraffa, M. D. Caro and F. Ferretti, eds, 'Cartographies of the Mind: Philosophy and Psychology in Intersection', Springer Netherlands, Dordrecht, pp. 65–80.
https://doi.org/10.1007/1-4020-5444-0_5

Mahlamäki, S. (2023), A Qualitative Investigation of Binaural Spatial Music in Virtual Reality, *in* 'Proceedings of the 20th Sound and Music Computing Conference', Stockholm, Sweden, pp. 304–311.
https://www.researchgate.net/publication/374234051

Makransky, G., Terkildsen, T. S. and Mayer, R. E. (2019), 'Adding immersive virtual reality to a science lab simulation causes more presence but less learning', *Learning and Instruction* **60**, 225–236.
https://www.sciencedirect.com/science/article/pii/S0959475217303274

Malpica, S., Serrano, A., Allue, M., Bedia, M. G. and Masia, B. (2020), 'Crossmodal perception in virtual reality', *Multimedia Tools and Applications* **79**(5-6), 3311–3331.
http://link.springer.com/10.1007/s11042-019-7331-z

Mangalam, M., Oruganti, S., Buckingham, G. and Borst, C. W. (2024), 'Enhancing hand-object interactions in virtual reality for precision manual tasks', *Virtual Reality* **28**(4), 166.
https://doi.org/10.1007/s10055-024-01055-3

Mantovani, F. (2001), VR Learning: Potential and Challenges for the Use of 3D, *in* 'Towards cyberpsychology: mind, cognition, and society in the Internet age', Vol. 2, Ios Press, pp. 207–225.

Marcus, A. and Rosenzweig, E., eds (2020), *Design, User Experience, and Usability. Interaction Design: 9th International Conference, DUXU 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I*, Vol. 12200 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham.
http://link.springer.com/10.1007/978-3-030-49713-2

Marks, L. E. (1987), 'On cross-modal similarity: Auditory–visual interactions in speeded discrimination', *Journal of Experimental Psychology: Human Perception and Performance* **13**(3), 384–394. Place: US.
https://doi.org/10.1037/0096-1523.13.3.384

Marougkas, A., Troussas, C., Krouska, A. and Sgouropoulou, C. (2024), 'How personalized and effective is immersive virtual reality in education? A systematic literature review for the last decade', *Multimedia Tools and Applications* **83**(6), 18185–18233.
https://doi.org/10.1007/s11042-023-15986-7

Martin, D., Malpica, S., Gutierrez, D., Masia, B. and Serrano, A. (2022), 'Multimodality in VR: A survey', *ACM Computing Surveys* p. 3508361.
https://dl.acm.org/doi/10.1145/3508361

Martino, G. and Marks, L. E. (1999), 'Perceptual and Linguistic Interactions in Speeded Classification: Tests of the Semantic Coding Hypothesis', *Perception* **28**(7), 903–923.
https://doi.org/10.1068/p2866

Martino, G. and Marks, L. E. (2000), 'Cross-Modal Interaction between Vision and Touch: The Role of Synesthetic Correspondence', *Perception* **29**(6), 745–754.
http://journals.sagepub.com/doi/10.1068/p2984

Martino, G. and Marks, L. E. (2001), 'Synesthesia: Strong and Weak', *Current Directions in Psychological Science* **10**(2), 61–65.
https://doi.org/10.1111/1467-8721.00116

Marto, A., Gonçalves, A., Melo, M. and Bessa, M. (2022), 'A survey of multisensory VR and AR applications for cultural heritage', *Computers & Graphics* **102**, 426–440.
https://linkinghub.elsevier.com/retrieve/pii/S0097849321002120

Marucci, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E. F., Babiloni, F., Betti, V. and Aricò, P. (2021), 'The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence', *Scientific Reports* **11**(1), 4831.
https://www.nature.com/articles/s41598-021-84196-8

Mattila, O., Korhonen, A., Pöyry, E., Hauru, K., Holopainen, J. and Parvinen, P. (2020), 'Restoration in a virtual reality forest environment', *Computers in Human Behavior* **107**, 106295.
https://www.sciencedirect.com/science/article/pii/S0747563220300509

Mayer, R. E. (2005), Cognitive Theory of Multimedia Learning, *in* R. Mayer, ed., 'The Cambridge Handbook of Multimedia Learning', Cambridge Handbooks in Psychology, Cambridge University Press, Cambridge, pp. 31–48.
https://www.cambridge.org/core/books/cambridge-handbook-of-multimedia-learning/cognitive-theory-of-multimedia-learning/A49922ACB5BC6A37DDCCE4131AC217E5

Mayer, R. E. (2023), 'Nine Ways to Reduce Cognitive Load in Multimedia Learning'.
https://www.tandfonline.com/doi/epdf/10.1207/S15326985EP3801_6?needAccess=true

McMahan, R. P., Gorton, D., Gresock, J., McConnell, W. and Bowman, D. A. (2006), Separating the effects of level of immersion and 3D interaction techniques, *in* 'Proceedings of the ACM symposium on Virtual reality software and technology', VRST '06, Association for Computing Machinery, New York, NY, USA, pp. 108–111.
https://dl.acm.org/doi/10.1145/1180495.1180518

Melara, R. D. (1989), 'Dimensional interaction between color and pitch', *Journal of Experimental Psychology: Human Perception and Performance* **15**(1), 69–79. Place: US.
https://doi.org/10.1037/0096-1523.15.1.69

Melara, R. D. and O'Brien, T. P. (1987), 'Interaction between synesthetically corresponding dimensions', *Journal of Experimental Psychology: General* **116**(4), 323–336. Place: US.
https://doi.org/10.1037/0096-3445.116.4.323

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W. and Davis, T. J. (2014), 'Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis', *Computers & Education* **70**, 29–40.
https://www.sciencedirect.com/science/article/pii/S0360131513002108

Merleau-Ponty, M. (2002), *Phenomenology of perception: an introduction*, Routledge, London. OCLC: 80428978.

Merter, S. (2017), 'Synesthetic Approach in the Design Process for Enhanced Creativity and Multisensory Experiences', *The Design Journal* **20**(sup1), S4519–S4528.
https://www.tandfonline.com/doi/full/10.1080/14606925.2017.1352948

Mevlevioğlu, D., Tabirca, S. and Murphy, D. (2023), Emotional Virtual Reality Stroop Task, *in* '2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)', IEEE, Atlanta, GA, USA, pp. 172–177.
https://ieeexplore.ieee.org/document/10150350/

Mine, M. R. (1995), Virtual Environment Interaction Techniques, Technical report, University of North Carolina.
https://dl.acm.org/doi/abs/10.5555/897820#core-collateral-info

Mine, M. R., Brooks, F. P. and Sequin, C. H. (1997), Moving objects in space: exploiting proprioception in virtual-environment interaction, *in* 'Proceedings of the 24th annual conference on Computer graphics and interactive techniques', SIGGRAPH '97, ACM Press/Addison-Wesley Publishing Co., USA, pp. 19–26.
https://dl.acm.org/doi/10.1145/258734.258747

Molero, D., Schez-Sobrino, S., Vallejo, D., Glez-Morcillo, C. and Albusac, J. (2021), 'A novel approach to learning music and piano based on mixed reality and gamification', *Multimedia Tools and Applications* **80**(1), 165–186.
https://doi.org/10.1007/s11042-020-09678-9

Mono vs Stereo Audio: Understanding the Differences (n.d.).
https://www.usemogul.com/post/mono-vs-stereo-audio-understanding-the-differences

Moon, K. and Blackman, D. (2017), 'A guide to ontology, epistemology, and philosophical perspectives for interdisciplinary researchers'.
https://i2insights.org/2017/05/02/philosophy-for-interdisciplinarity/

Morgan, D. L. (2007), 'Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods', *Journal of Mixed Methods Research* **1**(1), 48–76.
https://doi.org/10.1177/2345678906292462

Morrison, N. M. V., Burnham, D. and Morrison, B. W. (2015), 'Cognitive Load in Cross-Modal Dual-Task Processing', *Applied Cognitive Psychology* **29**(3), 436–444. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.3122.
https://onlinelibrary.wiley.com/doi/abs/10.1002/acp.3122

Murgia, M. D. (2015), 'Teaching Synaesthesia as a Gateway to Creativity', *Exchanges: The Interdisciplinary Research Journal* **2**(2), 305–313.
https://doi.org/10.31273/eirj.v2i2.118

Mutlu-Bayraktar, D., Cosgun, V. and Altan, T. (2019), 'Cognitive load in multimedia learning environments: A systematic review', *Computers & Education* **141**, 103618.
https://www.sciencedirect.com/science/article/pii/S036013151930171X

Nanjappan, V., Uunila, A., Vaulanen, J., Välimaa, J. and Georgiev, G. V. (2023), 'EFFECTS OF IMMERSIVE VIRTUAL REALITY IN ENHANCING CREATIVITY', *Proceedings of the Design Society* **3**, 1585–1594.
https://www.cambridge.org/core/journals/proceedings-of-the-design-society/article/effects-of-immersive-virtual-reality-in-enhancing-creativity/FEC35A450AE59B599D43346C5B562684

Ontiveros-Hernández, N. J., Pérez-Ramírez, M. and Hernández, Y. (2013), 'Virtual Reality and Affective Computing for Improving Learning', *Research in Computing Science* **65**(1), 121–131.
https://www.researchgate.net/publication/339203195_Virtual_Reality_and_Affective_Computing_for_Improving_Learning

Onwuegbuzie, A. J. and Collins, K. M. T. (2007), 'A Typology of Mixed Methods Sampling Designs in Social Science Research', *Qualitative Report* **12**(2), 281–316. ERIC Number: EJ800183.
https://eric.ed.gov/?id=EJ800183

O'Toole, P. (2021), Understanding Personalised Auditory-Visual Associations in Multi-Modal Interactions, *in* 'Proceedings of the 2021 International Conference on Multimodal Interaction', ACM, Montréal QC Canada, pp. 812–816.
https://dl.acm.org/doi/10.1145/3462244.3481277

O'Toole, P., Glowinski, D. and Mancini, M. (2019), Understanding Chromaesthesia by Strengthening Auditory -Visual-Emotional Associations, *in* '2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)', IEEE, Cambridge, United Kingdom, pp. 1–7.
https://ieeexplore.ieee.org/document/8925465/

O'Toole, P., Glowinski, D., Pitt, I. and Mancini, M. (2021), When Emotions are Triggered by Single Musical Notes: Revealing the Underlying Factors of Auditory-Emotion Associations, *in* 'Companion Publication of the 2021 International Conference on Multimodal Interaction', ACM, Montreal QC Canada, pp. 291–298.
https://dl.acm.org/doi/10.1145/3461615.3485419

O'Toole, P., Mancini, M., Pitt, I. and Maye, L. (2024), Chromaesthetica: A Virtual Reality App for Learning Musical Theory Using Sound-Colour Cross-Modal Associations, *in* 'Proceedings of the International Conference on Mobile and Ubiquitous Multimedia', MUM '24, Association for Computing Machinery, New York, NY, USA, pp. 513–515.
https://dl.acm.org/doi/10.1145/3701571.3703369

Ou, Y. (2023), 'Multimodal Music Teaching Mode Based on Human-computer Interaction Technology', *International Journal of Educational Curriculum Management and Research* **4**(3).
http://scholar-press.com/papers/1657

Oxenham, A. J. (2012), 'Pitch Perception', *The Journal of Neuroscience* **32**(39), 13335–13338.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3481156/

O'Toole, P., Mancini, M., Pitt, I. and Maye, L. (2025), 'Investigating musical pitch-colour associations in immersive virtual reality environments', *Virtual Reality* **29**(2), 88.
https://doi.org/10.1007/s10055-025-01163-8

Paas, F. G. W. C. (1992), 'Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach.', *Journal of Educational*

Psychology **84**(4), 429–434.
https://doi.org/10.1037/0022-0663.84.4.429

Padhi, D. R. and Katoch, S. (2021), How Real is Virtual reality—An Immersion, Interaction and Embodiedness Study, *in* A. Chakrabarti, R. Poovaiah, P. Bokil and V. Kant, eds, 'Design for Tomorrow—Volume 2', Springer, Singapore, pp. 805–817.
https://doi.org/10.1007/978-981-16-0119-4_65

Padmanaban, N. and Molner, K. (2016), 'Explorations in Spatial Audio and Perception for Virtual Reality'.
https://web.stanford.edu/class/ee267/Spring2016/report_padmanaban_molner.pdf

Pallavicini, F. and Pepe, A. (2019), Comparing Player Experience in Video Games Played in Virtual Reality or on Desktop Displays: Immersion, Flow, and Positive Emotions, *in* 'Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts', ACM, Barcelona Spain, pp. 195–210.
https://dl.acm.org/doi/10.1145/3341215.3355736

Pallavicini, F., Pepe, A. and Minissi, M. E. (2019), 'Gaming in Virtual Reality: What Changes in Terms of Usability, Emotional Response and Sense of Presence Compared to Non-Immersive Video Games?', *Simulation & Gaming* **50**(2), 136–159.
https://doi.org/10.1177/1046878119831420

Palmer, S. E., Schloss, K. B., Xu, Z. and Prado-Leon, L. R. (2013), 'Music-color associations are mediated by emotion', *Proceedings of the National Academy of Sciences* **110**(22), 8836–8841.
http://www.pnas.org/cgi/doi/10.1073/pnas.1212562110

Papert, S. and Harel, I. (1991), Situating Constructionism, *in* 'Constructionism', Ablex Publishing, pp. 1–11.
https://hcs64.com/teaching%20CS/papert-situating_constructionism.pdf

Parise, C. V. and Spence, C. (2009), "When Birds of a Feather Flock Together': Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes', *PLOS ONE* **4**(5), e5664.
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0005664

Parise, C. V. and Spence, C. (2012), 'Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test', *Experimental Brain Research* **220**(3-4), 319–333.
http://link.springer.com/10.1007/s00221-012-3140-6

Parseihian, G. and Katz, B. F. G. (2012), 'Rapid head-related transfer function adaptation using a virtual auditory environment', *The Journal of the Acoustical Society of America* **131**(4), 2948–2957.
https://asa.scitation.org/doi/abs/10.1121/1.3687448

Parsons, T. D. and Phillips, A. S. (2016), 'Virtual reality for psychological assessment in clinical practice.', *Practice Innovations* **1**(3), 197–217.
https://doi.org/10.1037/pri0000028

Patching, G. R. and Quinlan, P. T. (2002), 'Garner and congruence effects in the speeded classification of bimodal signals', *Journal of Experimental Psychology: Human Perception and Performance* **28**(4), 755–775. Place: US.
https://doi.org/10.1037/0096-1523.28.4.755

Pekrun, R. and Stephens, E. J. (2012), Academic emotions, *in* 'APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors', APA handbooks in psychology®, American Psychological Association, Washington, DC, US, pp. 3–31.
https://doi.org/10.1037/13274-001

Pellas, N., Kazanidis, I. and Palaigeorgiou, G. (2020), 'A systematic literature review of mixed reality environments in K-12 education', *Education and Information Technologies* **25**(4), 2481–2520.
https://doi.org/10.1007/s10639-019-10076-4

Pereira, D., Oliveira, V., Vilaça, J. L., Carvalho, V. and Duque, D. (2023), 'Measuring the Precision of the Oculus Quest 2's Handheld Controllers', *Actuators* **12**(6), 257. Number: 6.
https://www.mdpi.com/2076-0825/12/6/257

Pessoa, L. (2008), 'On the relationship between emotion and cognition', *Nature Reviews Neuroscience* **9**(2), 148–158.
https://www.nature.com/articles/nrn2317

Plutchik, R. (2001), 'The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice', *American Scientist* **89**(4), 344–350.
https://www.jstor.org/stable/27857503

Poeschl, S., Wall, K. and Doering, N. (2013), Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence, *in* '2013 IEEE Virtual Reality (VR)', IEEE, Lake Buena Vista, FL, pp. 129–130.
http://ieeexplore.ieee.org/document/6549396/

Poguntke, R., Wirth, M. and Gradl, S. (2019), Same Same but Different: Exploring the Effects of the Stroop Color Word Test in Virtual Reality, *in* D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler and P. Zaphiris, eds, 'Human-Computer Interaction – INTERACT 2019', Vol. 11747, Springer International Publishing, Cham, pp. 699–708. Series Title: Lecture Notes in Computer Science.
http://link.springer.com/10.1007/978-3-030-29384-0_42

Potter, T., Cvetković, Z. and De Sena, E. (2022), 'On the Relative Importance of Visual and Spatial Audio Rendering on VR Immersion', *Frontiers in Signal Processing* **2**.
https://www.frontiersin.org/articles/10.3389/frsip.2022.904866

Pouw, W. T. J. L., van Gog, T. and Paas, F. (2014), 'An Embedded and Embodied Cognition Review of Instructional Manipulatives', *Educational Psychology Review* **26**(1), 51–72.
https://doi.org/10.1007/s10648-014-9255-5

Ppali, S., Lalioti, V., Branch, B., Ang, C. S., Thomas, A. J., Wohl, B. S. and Covaci, A. (2022), Keep the VRhythm going: A musician-centred study investigating how Virtual Reality can support creative musical practice, *in* 'Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems', CHI '22, Association for Computing Machinery, New York, NY, USA, pp. 1–19.
https://doi.org/10.1145/3491102.3501922

Ramachandran, V. S. and Hubbard, E. M. (2001), 'Psychophysical investigations into the neural basis of synaesthesia', *Proceedings of the Royal Society of London. Series B: Biological Sciences* **268**(1470), 979–983.
https://royalsocietypublishing.org/doi/10.1098/rspb.2000.1576

Ramachandran, V. S. and Hubbard, E. M. (2003), 'Hearing Colors, Tasting Shapes', *Scientific American* **288**(5), 52–59.
http://www.jstor.org.ucc.idm.oclc.org/stable/26060283

Ramírez-Sanz, J. M., Pena-Alonso, H. M., Serrano-Mamolar, A., Arnaiz-Gonzalez, A. and Bustillo, A. (2023), Detection of Stress Stimuli in Learning Contexts of iVR Environments, *in* L. T. De Paolis, P. Arpaia and M. Sacco, eds, 'Extended Reality', Vol. 14219, Springer Nature Switzerland, Cham, pp. 427–440. Series Title: Lecture Notes in Computer Science.
https://link.springer.com/10.1007/978-3-031-43404-4_29

Regenbrecht, H. and Schubert, T. (2002), 'Real and Illusory Interactions Enhance Presence in Virtual Environments', *PRESENCE: Teleoperators & Virtual Environments* **11**(4), 425–434.
https://ucc.idm.oclc.org/login?URL=https://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=7188709&site=ehost-live

Rettinger, M., Jiang, X., Yang, J. and Rigoll, G. (2024), Visual Complexity in VR: Implications for Cognitive Load, *in* '2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)', pp. 783–784.
https://ieeexplore.ieee.org/document/10536293

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C. and Alcañiz, M. (2007), 'Affective Interactions Using Virtual Reality: The Link between Presence and Emotions', *CyberPsychology & Behavior* **10**(1), 45–56.
http://www.liebertpub.com/doi/10.1089/cpb.2006.9993

Rucsanda, M. D. (2017), Synaesthesia and artistic experimentation in Alexandr Scriabin's works, PhD thesis, University of Braşov.
https://www.ceeol.com/search/article-detail?id=617278

Rudenko, S. and Cabral, J. P. (2019), Synaesthesia: How can it be used to enhance the audio-visual perception of music and multisensory design in digitally enhanced

environments?, *in* 'Workshop on Speech, Music and Mind (SMM 2019)', ISCA, pp. 60–64.
https://www.isca-speech.org/archive/smm_2019/rudenko19_smm.html

Rudenko, S. and Haahr, M. (2022), Psychogeography with Jack B. Yeats Art Sounding Gallery: Augmented Reality Locative Experience for Blind People, *in* 'Proceedings of the 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022)'.
https://openaccess.cms-conferences.org/publications/book/978-1-958651-13-1/article/978-1-958651-13-1_3

Rudenko, S., McDonnell, M., Layden, T. and Haahr, M. (2022), Exploring Classical Music Narratives Through Multimodality in AR/VR Experiences, *in* M. Vosmeer and L. Holloway-Attaway, eds, 'Interactive Storytelling', Vol. 13762, Springer International Publishing, Cham, pp. 395–404. Series Title: Lecture Notes in Computer Science.
https://link.springer.com/10.1007/978-3-031-22298-6_24

Sagiv, N. and Ward, J. (2006), Chapter 15 Crossmodal interactions: lessons from synesthesia, *in* 'Progress in Brain Research', Vol. 155, Elsevier, pp. 259–271.
https://linkinghub.elsevier.com/retrieve/pii/S0079612306550150

Salatino, A., Zavattaro, C., Gammeri, R., Cirillo, E., Piatti, M. L., Pyasik, M., Serra, H., Pia, L., Geminiani, G. and Ricci, R. (2023), 'Virtual reality rehabilitation for unilateral spatial neglect: A systematic review of immersive, semi-immersive and non-immersive techniques', *Neuroscience & Biobehavioral Reviews* **152**, 105248.
https://www.sciencedirect.com/science/article/pii/S0149763423002178

Sargunam, S. P. and Ragan, E. D. (2018), Evaluating joystick control for view rotation in virtual reality with continuous turning, discrete turning, and field-of-view reduction, *in* 'Proceedings of the 3rd International Workshop on Interactive and Spatial Computing', IWISC '18, Association for Computing Machinery, New York, NY, USA, pp. 74–79.
https://dl.acm.org/doi/10.1145/3191801.3191815

Schrader, C. and Bastiaens, T. J. (2012), 'The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games', *Computers in Human Behavior* **28**(2), 648–658.
https://www.sciencedirect.com/science/article/pii/S074756321100255X

Schubert, T., Friedmann, F. and Regenbrecht, H. (2001), 'The Experience of Presence: Factor Analytic Insights', *Presence: Teleoperators and Virtual Environments* **10**(3), 266–281.
https://direct.mit.edu/pvar/article/10/3/266-281/18310

Schwind, V., Knierim, P., Haas, N. and Henze, N. (2019), Using Presence Questionnaires in Virtual Reality, *in* 'Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems', ACM, Glasgow Scotland Uk, pp. 1–12.
https://dl.acm.org/doi/10.1145/3290605.3300590

Shams, L. and Seitz, A. R. (2008), 'Benefits of multisensory learning', *Trends in Cognitive Sciences* **12**(11), 411–417.
https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(08)00218-0

Shirokova, M. (2021), Can You Hear Your Drawing? - Creating synesthetic experiences through an audio-visual web interaction, PhD thesis, OCAD University - Design in Digital Futures, Toronto, Ontario, Canada.
https://openresearch.ocadu.ca/id/eprint/3330/

Silva Bastos, A., Gomes, R. F., dos Santos, C. C. and Maia, J. G. R. (2018), 'Synesthesia: A Study on Immersive Features of Electronic Games', *Journal on Interactive Systems* **9**(2), 1.
https://sol.sbc.org.br/journals/index.php/jis/article/view/700

Sinlapanuntakul, W., Derby, J. L. and Chaparro, B. S. (2022), 'Understanding the Effects of Mixed Reality on Video Game Satisfaction, Enjoyment, and Performance', *Simulation & Gaming* **53**(3), 237–252.
http://journals.sagepub.com/doi/10.1177/10468781221094473

Slater, M., Steed, A., McCarthy, J. and Maringelli, F. (1998), 'The Influence of Body Movement on Subjective Presence in Virtual Environments', *Human Factors: The Journal of the Human Factors and Ergonomics Society* **40**(3), 469–477.
http://journals.sagepub.com/doi/10.1518/001872098779591368

Slater, M. and Wilbur, S. (1997), 'A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments', *Presence: Teleoperators and Virtual Environments* **6**(6), 603–616.
https://direct.mit.edu/pvar/article/6/6/603-616/18157

Smith, J. A., Larkin, M. and Flowers, P. (2021), 'Interpretative Phenomenological Analysis : Theory, Method and Research', pp. 1–100.
https://www.torrossa.com/it/resources/an/5282221

Soegaard, M. (2025), 'Revolutionize UX Design with VR Experiences'.
https://www.interaction-design.org/literature/article/ixdf-ux-design-for-virtual-reality-course

Soraghan, S. (2014), Animating Timbre - A User Study, *in* 'ICMC|SMC', Athens, Greece.
https://www.icmc14-smc14.net/images/proceedings/OS8-B04-AnimatingTimbre.pdf

Spector, F. and Maurer, D. (2009), 'Synesthesia: A new approach to understanding the development of perception.', *Developmental Psychology* **45**(1), 175–189.
http://doi.apa.org/getdoi.cfm?doi=10.1037/a0014171

Spence, C. (2007), 'Audiovisual multisensory integration', *Acoustical Science and Technology* **28**(2), 61–70.
https://doi.org/10.1250/ast.28.61

Spence, C. (2011), 'Crossmodal correspondences: A tutorial review', *Attention, Perception, & Psychophysics* **73**(4), 971–995.
https://doi.org/10.3758/s13414-010-0073-7

Stein, B. E. and Stanford, T. R. (2008), 'Multisensory integration: current issues from the perspective of the single neuron', *Nature Reviews Neuroscience* **9**(4), 255–266.
https://www.nature.com/articles/nrn2331

Stevenson, R. A., Zemtsov, R. K. and Wallace, M. T. (2012), 'Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions', *Journal of Experimental Psychology: Human Perception and Performance* **38**(6), 1517–1529. Place: US.
https://doi.org/10.1037/a0027339

Stone, W. B. (2017), Psychometric Evaluation of the Simulator Sickness Questionnaire as a Measure of Cybersickness, PhD thesis, Iowa State University, Iowa State University.
https://www.proquest.com/openview/c2d59f535cdea2e705131a3b052a1b05/1?pq-origsite=gscholar&cbl=18750

Sun, X., Li, X., Ji, L., Han, F., Wang, H., Liu, Y., Chen, Y., Lou, Z. and Li, Z. (2018), 'An extended research of crossmodal correspondence between color and sound in psychology and cognitive ergonomics', *PeerJ* **6**, e4443.
https://peerj.com/articles/4443

Sun, Y., Yao, L. and Fu, Q. (2024), 'Crossmodal Correspondence Mediates Crossmodal Transfer from Visual to Auditory Stimuli in Category Learning', *Journal of Intelligence* **12**(9), 80. Number: 9.
https://www.mdpi.com/2079-3200/12/9/80

Sweller, J. (2011), CHAPTER TWO - Cognitive Load Theory, *in* J. P. Mestre and B. H. Ross, eds, 'Psychology of Learning and Motivation', Vol. 55, Academic Press, pp. 37–76.
https://www.sciencedirect.com/science/article/pii/B9780123876911000028

Sweller, J. (2020), 'Cognitive load theory and educational technology', *Educational Technology Research and Development* **68**(1), 1–16.
https://doi.org/10.1007/s11423-019-09701-3

Szucs, V., Kovacs, B. and Tasnadi, B. (2018), Music for seeing – visualization of sounds, *in* '2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)', pp. 000123–000128.
https://ieeexplore.ieee.org/abstract/document/8639871

Timoney, J., Faghih, B., Gibney, A., Korady, B. and Young, G. (2018), SingingBlocks: Considerations for a Virtual Reality Game to create chords and progressions, *in* 'Computer Simulation of Musical Creativity (CSMC)', pp. 1–13.
https://mural.maynoothuniversity.ie/id/eprint/13320/

Tomasetti, M. and Turchet, L. (2024), 'Handheld controller-based locomotion in Virtual Reality as an approach to interactive music composition: insights from composers' preferences', *Digital Creativity* **0**(0), 1–25. _eprint: https://doi.org/10.1080/14626268.2024.2386353.
https://doi.org/10.1080/14626268.2024.2386353

Trujano, F., Khan, M. and Maes, P. (2018), ARPiano Efficient Music Learning Using Augmented Reality, *in* T.-T. Wu, Y.-M. Huang, R. Shadiev, L. Lin and A. I. Starčič, eds, 'Innovative Technologies and Learning', Vol. 11003, Springer International Publishing, Cham, pp. 3–17. Series Title: Lecture Notes in Computer Science.
http://link.springer.com/10.1007/978-3-319-99737-7_1

Tsang, T. and Schloss, K. B. (2010), 'Associations between Color and Music are Mediated by Emotion and Influenced by Tempo: (525772013-006)'. Pages: 82-93.
https://bpb-us-w2.wpmucdn.com/campuspress.yale.edu/dist/a/1215/files/2015/11/2010_YRURP-Second-Issue-1c48m7l.pdf

Tutescu, A.-S. (2022), Flow, intuitive interaction design and music production in an audio-first VR experience, Master's thesis, HOCHSCHULE FÜR ANGEWANDTE WISSENSCHAFTEN HAMBURG, Hamburg.
https://hdl.handle.net/20.500.12738/16140

Tyng, C. M., Amin, H. U., Saad, M. N. M. and Malik, A. S. (2017), 'The Influences of Emotion on Learning and Memory', *Frontiers in Psychology* **8**.
https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.01454/full

Usoh, M., Catena, E., Arman, S. and Slater, M. (2000), 'Using Presence Questionnaires in Reality', *Presence: Teleoperators and Virtual Environments* **9**(5), 497–503.
https://direct.mit.edu/pvar/article/9/5/497-503/18368

Valtchanov, D., Barton, K. R. and Ellard, C. (2010), 'Restorative Effects of Virtual Nature Settings', *Cyberpsychology, Behavior, and Social Networking* **13**(5), 503–512.
https://www.liebertpub.com/doi/full/10.1089/cyber.2009.0308

van Wassenhove, V., Grant, K. W. and Poeppel, D. (2007), 'Temporal window of integration in auditory-visual speech perception', *Neuropsychologia* **45**(3), 598–607.
https://www.sciencedirect.com/science/article/pii/S002839320600011X

Vasarainen, M., Paavola, S. and Vetoshkina, L. (2021), 'A Systematic Literature Review on Extended Reality: Virtual, Augmented and Mixed Reality in Working Life', *Internationa Journal of Virtual Reality* **21**(2), 1–28.
https://doi.org/10.20870/IJVR.2021.21.2.4620

Villena-Taranilla, R., Tirado-Olivares, S., Cózar-Gutiérrez, R. and González-Calero, J. A. (2022), 'Effects of virtual reality on learning outcomes in K-6 education: A meta-analysis', *Educational Research Review* **35**, 100434.
https://www.sciencedirect.com/science/article/pii/S1747938X22000033

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M. and Levitin, D. J. (2011), 'Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance', *Cognition* **118**(2), 157–170.
https://linkinghub.elsevier.com/retrieve/pii/S0010027710002817

Wager, A. (1999), 'The extra qualia problem: Synaesthesia and representationism', *Philosophical Psychology* **12**(3), 263–281. _eprint: https://doi.org/10.1080/095150899105756.
https://doi.org/10.1080/095150899105756

Wang, X., Shi, Y., Zhang, B. and Chiang, Y. (2019), 'The Influence of Forest Resting Environments on Stress Using Virtual Reality', *International Journal of Environmental Research and Public Health* **16**(18), 3263. Number: 18.
https://www.mdpi.com/1660-4601/16/18/3263

Ward, J. (2013), 'Synesthesia', *Annual Review of Psychology* **64**(Volume 64, 2013), 49–75.
https://www.annualreviews.org/content/journals/10.1146/annurev-psych-113011-143840

Ward, J., Huckstep, B. and Tsakanikos, E. (2006), 'Sound-Colour Synaesthesia: to What Extent Does it Use Cross-Modal Mechanisms Common to us All?', *Cortex* **42**(2), 264–280.
https://linkinghub.elsevier.com/retrieve/pii/S0010945208703526

Ward, J., Maciel, S., Rouw, R., Simner, J. and Root, N. (2025), 'Synaesthesia is linked to differences in music preference and musical sophistication and a distinctive pattern of sound-color associations', *Psychology of Music* **53**(3), 453–473.
https://journals.sagepub.com/doi/10.1177/03057356241250020

Ward, J. and Meijer, P. (2010), 'Visual experiences in the blind induced by an auditory sensory substitution device', *Consciousness and Cognition* **19**(1), 492–500.
https://www.sciencedirect.com/science/article/pii/S1053810009001718

Waterworth, J., Mantovani, F. and Riva, G. (2012), 'Special Issue: Presence and Interaction', *Interacting with Computers* **24**(4), 190–192.
https://doi.org/10.1016/j.intcom.2012.05.003

Weinel, J. (2021), Synaesthetic Audio-Visual Sound Toys in Virtual Reality, *in* 'Audio Mostly 2021', ACM, virtual/Trento Italy, pp. 135–138.
https://dl.acm.org/doi/10.1145/3478384.3478400

Whiteford, K. L., Schloss, K. B., Helwig, N. E. and Palmer, S. E. (2018), 'Color, Music, and Emotion: Bach to the Blues', *i-Perception* **9**(6), 1–27.
http://journals.sagepub.com/doi/10.1177/2041669518808535

Wiesing, M., Steinkönig, H. W., Vossel, S., Fink, G. R. and Weidner, R. (2022), Transferring paradigms from physical to virtual reality: Can reaction time effects be replicated in a virtual setting?, preprint, PsyArXiv.
https://osf.io/kmsnz

Williams, R., Gumtau, S. and Mackness, J. (2015), 'Synesthesia: From Cross-Modal to Modality-Free Learning and Knowledge', *Leonardo* **48**(1), 48–54.
http://www.mitpressjournals.org/doi/10.1162/LEON_a_00937

Wilson, M. (2002), 'Six views of embodied cognition', *Psychonomic Bulletin & Review* **9**(4), 625–636.
https://doi.org/10.3758/BF03196322

Wilson, P. T., Kalescky, W., MacLaughlin, A. and Williams, B. (2016), VR locomotion: walking > walking in place > arm swinging, *in* 'Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry - Volume 1', ACM, Zhuhai China, pp. 243–249.
https://dl.acm.org/doi/10.1145/3013971.3014010

Witmer, B. G. and Singer, M. J. (1998), 'Measuring Presence in Virtual Environments: A Presence Questionnaire', *Presence: Teleoperators and Virtual Environments* **7**(3), 225–240.
https://direct.mit.edu/pvar/article/7/3/225-240/92643

Wong, Y. S., Wang, I. T. and Hashim, M. N. (2023), 'Examining Scaffolding Process in One-to-One Piano Lessons for Young Beginners: An Observational Analysis', *Malaysian Journal of Music* **12**(1), 66–79. Number: 1.
https://ejournal.upsi.edu.my/index.php/MJM/article/view/7913

Wright, V. and Kefalidou, G. (2021), Can you hear the Colour? Towards a Synaesthetic and Multimodal Design Approach in Virtual Worlds, *in* 'Proceedings of the BCS', British HCI Conference, London.
http://dx.doi.org/10.14236/ewic/HCI2021.4

Wu, D., Courtney, C. G., Lance, B. J., Narayanan, S. S., Dawson, M. E., Oie, K. S. and Parsons, T. D. (2010), 'Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task', *IEEE Transactions on Affective Computing* **1**(2), 109–118.
http://ieeexplore.ieee.org/document/5611481/

Wu, F., Bailey, G. S., Stoffregen, T. and Suma Rosenberg, E. (2021), Don't Block the Ground: Reducing Discomfort in Virtual Reality with an Asymmetric Field-of-View Restrictor, *in* 'Proceedings of the 2021 ACM Symposium on Spatial User Interaction', SUI '21, Association for Computing Machinery, New York, NY, USA, pp. 1–10.
https://dl.acm.org/doi/10.1145/3485279.3485284

Yau, D. and Mccrindle, R. (2007), 'MusiCam: an instrument to demonstrate chromaphonic synesthesia', *Digital Creativity* **18**(2), 121–127.
http://www.tandfonline.com/doi/abs/10.1080/14626260701401510

Young, G. W., O'Dwyer, N., Johnson, N. and Smolic, A. (2025), 'Extended reality in higher education: A case study in film, theatre, and performance practice', *Arts and Humanities in Higher Education* p. 14740222251341403.
https://journals.sagepub.com/doi/10.1177/14740222251341403

Yu, C.-P., Lee, H.-Y. and Luo, X.-Y. (2018), 'The effect of virtual reality forest and urban environments on physiological and psychological responses', *Urban Forestry & Urban Greening* **35**, 106–114.
https://www.sciencedirect.com/science/article/pii/S1618866718301900

Yu, S., Liu, Q., Johnson-Glenberg, M. C., Han, M., Ma, J., Ba, S. and Wu, L. (2023), 'Promoting musical instrument learning in virtual reality environment: Effects of embodiment and visual cues', *Computers & Education* **198**, 104764.
https://www.sciencedirect.com/science/article/pii/S0360131523000416

Zhang, W., Samarasinghe, P. N., Chen, H. and Abhayapala, T. D. (2017), 'Surround by Sound: A Review of Spatial Audio Recording and Reproduction', *Applied Sciences* **7**(5), 532. Number: 5.
https://www.mdpi.com/2076-3417/7/5/532

Zhou, Y., Ji, S., Xu, T. and Wang, Z. (2018), 'Promoting Knowledge Construction: A Model for Using Virtual Reality Interaction to Enhance Learning', *Procedia Computer Science* **130**, 239–246.
https://linkinghub.elsevier.com/retrieve/pii/S1877050918303867

Çamcı, A. and Hamilton, R. (2020), 'Audio-first VR: New perspectives on musical experiences in virtual environments', *Journal of New Music Research* **49**(1), 1–7.
https://www.tandfonline.com/doi/full/10.1080/09298215.2019.1707234

Appendix A

Study One

A.1 Consent Form

Participant no:

- I, ... hereby agree to participate in this study to examine user's sound and colour association in three testing environments conducted by Patrick O'Toole under the supervision of Dr Ian Pitt and Dr Maurizio Mancini.
- I am 18 years old or older.
- I am participating voluntarily.
- The purpose of the study has been explained to me and I understand what is involved.
- I understand that I can withdraw from the study, without any repercussions.
- I understand that I can withdraw permission to use the data collected from me up to 14 days after my participation in the experiment, in which case the data will be deleted.
- I understand that my information will be kept confidential and my data will be anonymised.
- I give permission for my data to be used in any academic presentation or publication, including the PhD thesis.

Signed: .....
Date: .....

A.2 Information Sheet

This research investigates users associations between their auditory and visual senses using three testing methods. The current study is to conduct three tests for users associations between sound and colour stimuli, using a pen and paper, a standard computer and a Virtual Reality test. Should you accept to participate in this study, you will be asked to complete a short questionnaire for demographic information and to establish your previous VR experience. Once completed, We will start with one of the three tests with the sequenced randomised.

For the pen and paper test users will hear musical notes in random order and choose the colour they associate with the sound from the colour chart provided. The standard desktop test will include randomized musical notes play on the computer with the user choosing their colour association from a colour picker on the monitor. For the VR test, VR equipment will be placed on you by the researcher and you will be given a VR tutorial to familiarise you with using VR. Once the tutorial is over, you will begin the main VR experience where you will be asked to complete a similar sound-colour test to the previous two tests but in a VR environment. We will record the sound/colour association data in all three tests, and performance information e.g. head movement, in the VR test.

Participation is voluntary and you are free to withdraw from the experiment at any time. Participants can refuse to answer any questions in the questionnaires. During the VR task, if you decide to leave the task for any reason, pressing the exit button on the user interface will take you to an exit scene and the researcher will help you remove the equipment. Afterwards, you will be asked to complete a short questionnaire about your experience in the research.

All information collected from this experiment will be completely confidential, anonymised and stored on secure University College Cork systems. Each participant will be assigned a participant number, and the questionnaires and sound/colour data will be linked to this number indirectly. There will be no other way to connect the information to the participant. If you decide to withdraw from the research and do not want your data used, you can do so anytime up to 14 days from when the experiment took place.

The information collected will be used as part of a PhD research and will be used in academic presentations and publications.

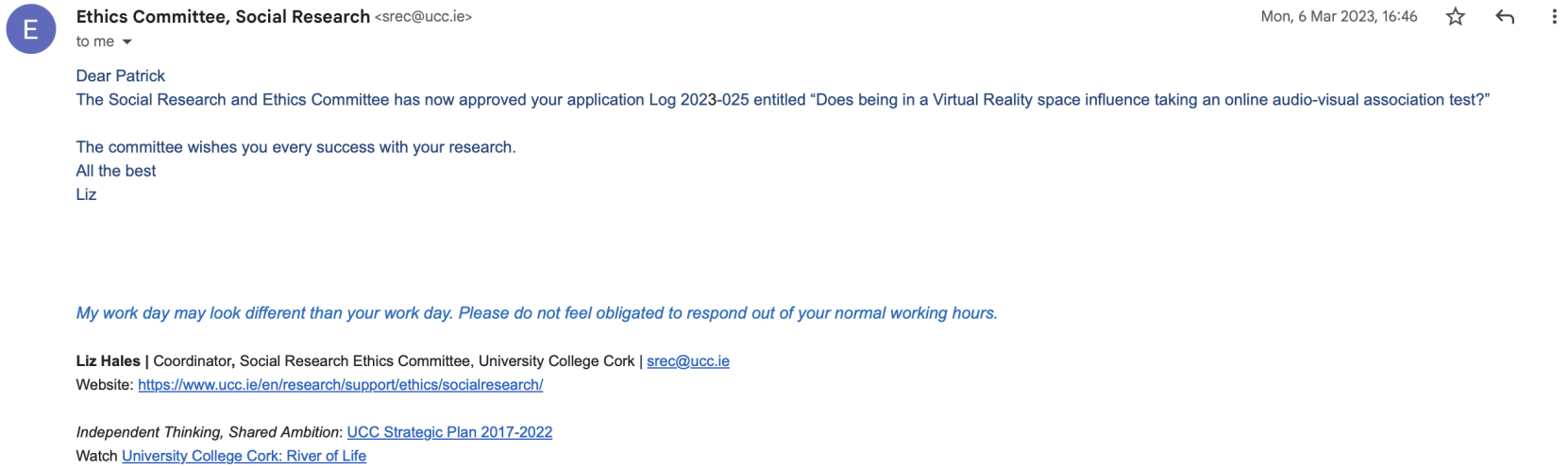
The study has been reviewed and received ethical approval from UCC's Social Research Ethics Committee.

If there is anything you'd like to know please contact me or my supervisors directly:

- Patrick O'Toole: patrick.otoole@umail.ucc.ie
- Dr. Ian Pitt: ianp@cs.ucc.ie
- Maurizio Mancini: m.mancini@di.uniroma1.it

Thank you for your time and if you agree to participate, please complete the study consent form.

A.3 Study One Ethics Approval



Ethics Committee, Social Research <srec@ucc.ie> Mon, 6 Mar 2023, 16:46

to me

Dear Patrick

The Social Research and Ethics Committee has now approved your application Log 2023-025 entitled "Does being in a Virtual Reality space influence taking an online audio-visual association test?"

The committee wishes you every success with your research.

All the best

Liz

My work day may look different than your work day. Please do not feel obligated to respond out of your normal working hours.

Liz Hales | Coordinator, Social Research Ethics Committee, University College Cork | srec@ucc.ie

Website: https://www.ucc.ie/en/research/support/ethics/socialresearch/

Independent Thinking, Shared Ambition: UCC Strategic Plan 2017-2022

Watch University College Cork: River of Life

Figure A.1: Approval from the University College Cork Ethics Committee for Study One presented in Chapter 4.

A.4 Pre-Study Questionnaire Items

1. **What age group do you associate with?**
    a. 18-24
    b. 25-34
    c. 35-44
    d. 45-54
    e. 55 or older
2. **What is your gender?**
    a. Male
    b. Female
    c. Other
3. **Have you ever experienced Virtual Reality before?**
    a. Never
    b. Only tried a couple of times
    c. 1-2 years of experience
    d. 2-5 years of experience
    e. 5+ years of experience
4. **How often do you experience motion sickness?**
    a. Never
    b. Rarely
    c. Sometimes
    d. Often
    e. Very Often
5. **Do you knowingly have synaesthesia?**
    (This is where you experience stimuli originally from either your auditory or visual senses also in the secondary sense — e.g., experience colours when you hear a sound.)
    a. Yes
    b. No
    c. I don't know

A.5 Post-Study Questionnaire Items

1. **How difficult did you find the task of associating colours to musical auditory stimuli?**
   (Please rate from 1 (very easy) to 5 (very hard)

   1. Very Easy
   2. Easy
   3. Ok
   4. Hard
   5. Very Hard

2. **Did you feel the experience of taking the test in a real life environment and in Virtual Reality was similar?**

   a. Yes

   b. No

3. **What, if any, aspect of the Virtual Reality experience did you like or prefer to the test in the real life environment?**

4. **Of the two tests that were conducted, the auditory-visual test on the standard desktop and the same test conducted in Virtual Reality, Which test do you feel like you were able to associate the sounds you heard with colours that you saw/perceived better?**

   a. Standard Desktop Test

   b. Virtual Reality Test

   c. Neither

   d. Both

5. **From the answer given in the last question, can you expand on why you feel you performed better it the option stated in your answer?**

6. **If you have any other insight on the study or using Virtual Reality, please let us know below.**

A.6 Evolution of Template Analysis Themes - Study One

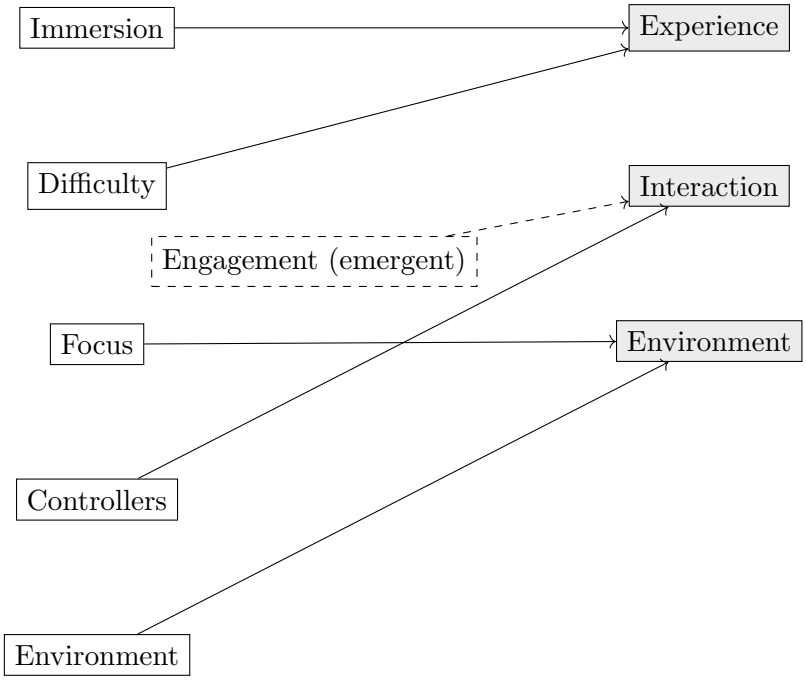


Figure A.2: Visualisation showing how the initial *"a priori"* codes (left side) were refined into final codes (right side) through iterative template analysis. Some initial codes were merged, others evolved into broader themes, and *"Engagement"* emerged as an additional code during the process before finally being grouped under *"Interaction"*.

Appendix B

Study Two

B.1 Consent Form

Participant no:

- I, ... hereby agree to participate in this usability study to test two VR environments conducted by Patrick O'Toole under the supervision of Dr Ian Pitt and Dr Maurizio Mancini.
- I am 18 years old or older.
- I am participating voluntarily.
- The purpose of the study has been explained to me and I understand what is involved.
- I understand that I can withdraw from the study, without any repercussions.
- I understand that I can withdraw permission to use the data collected from me up to 14 days after my participation in the experiment, in which case the data will be deleted.
- I understand that my information will be kept confidential and my data will be anonymised.
- I give permission for my data to be used in any academic presentation or publication, including the PhD thesis.

Signed: .....
Date: .....

B.2 Information Sheet

This research is designed to compare two virtual reality (VR) environments which will be presented using a VR head mounted display. Should you agree to participate in this study, you will be asked to enter the VR environment, answer some general questions and complete a short tutorial before starting the experience. Once you have completed the tutorial, you will start the first test chosen randomly from the two VR environment.

Both tests will involve you listening to musical notes played in the VR space, and interacting with objects, manipulating the colour and size of the object according to your perceived association with the sound playing. The VR equipment will be placed on you by the researcher, and you will be given time to answer the general questions and practice getting use to movement and interaction in the VR space. Once you finish the first test, you will see a questionnaire pop up on the screen which you can answer and once that is done, you will be given a break and then the second VR environment will start. The controls and instructions will be the same as in the first environment, and once that is done you will see the questionnaire pop up again.

Once the questions are answered the application will end and you can remove the headset. Participation is voluntary and you are free to withdraw from the experiment at any time. Participants can refuse to answer any questions in the questionnaires. During the VR task, if you decide to leave the task for any reason, let the researcher know and they will help you remove the equipment. Afterwards, you will be asked to take part in a short interview about your experience whilst taking part in the research.

All information collected from this experiment will be completely confidential, anonymised and stored on secure University College Cork systems. Each participant will be assigned a participant number, and the questionnaires and sound/colour data will be linked to this number indirectly. There will be no other way to connect the information to the participant. If you decide to withdraw from the research and do not want your data used, you can do so anytime up to 14 days from when the experiment took place.

The information collected will be used as part of a PhD research project and will be used in academic presentations and publications. Some potential health issues associated with using head-mounted displays (HMDs) include:

- Eye strain: HMDs can cause eye strain, fatigue, and dry eyes, as the eyes have to focus on a close-up image for an extended period of time.

- Headaches: Some people may experience headaches or migraines due to the weight of the HMD or the strain on the eyes.

- Motion sickness: HMDs can cause motion sickness, as the brain receives conflicting signals from the eyes and inner ear.

If you feel uncomfortable or sick during the Virtual Reality experience as part of this study, please inform the researcher, and they will aid you. Once the researcher is informed, they will help you out of the VR headset and provide you with a comfortable seat; some water; and monitor your condition.

If you feel unwell after you finish the study, here are the contact details for the student health centre.

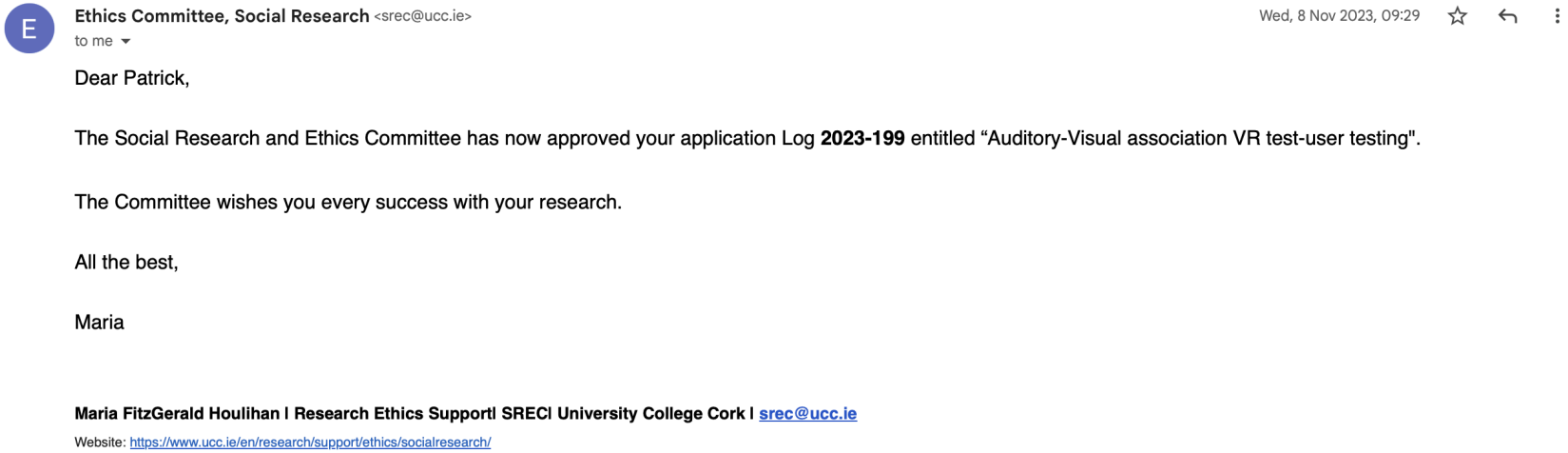
Tel: +353 (0)21 4902311
Address: Student Health Centre,
Crow's Nest, Victoria Cross Rd.
Cork, T12 HXW4

The study has been reviewed and has received ethical approval from UCC's Social Research Ethics Committee. If there is anything you'd like to know please contact me or my supervisors directly:

- Patrick O'Toole: patrick.otoole@umail.ucc.ie
- Dr. Ian Pitt: ianp@cs.ucc.ie
- Maurizio Mancini: m.mancini@di.uniroma1.it

Thank you for your time, please complete the study consent form to confirm your participation in the study.

B.3 Study Two Ethics Approval



Ethics Committee, Social Research <srec@ucc.ie>
to me

Wed, 8 Nov 2023, 09:29

Dear Patrick,

The Social Research and Ethics Committee has now approved your application Log **2023-199** entitled “Auditory-Visual association VR test-user testing".

The Committee wishes you every success with your research.

All the best,

Maria

Maria FitzGerald Houlihan I Research Ethics SupportI SRECI University College Cork I srec@ucc.ie
Website: https://www.ucc.ie/en/research/support/ethics/socialresearch/

Figure B.1: Approval from the University College Cork Ethics Committee for Study Two presented in Chapter 5.

B.4 Pre-Study Questionnaire Items

1. **Gender?**
    - Male
    - Female
    - Non-Binary
    - Other
2. **Age Group?**
    - 18-25
    - 26-35
    - 36-45
    - 46-55
    - 56-65
    - 66+
3. **Hand Preference?**
    - Right-handed
    - Left-handed
4. **Musical Instrument?**
    - Piano
    - Guitar
    - Violin
    - Percussion
    - Singer
    - Other
5. **Creative Hobbies?**
    - Art
    - Music
    - Writing
    - Film Making
    - Acting
    - Photography
6. **What is your favourite colour?**
7. **What is your least favourite colour?**

B.5 Post-Study Questionnaire Items

Below is a list of questions that were asked participants after completing the study, which were used as part of the template analysis.

1. **Please state your User ID.**
2. **Please indicate your preferred environment for each of the listed features below.**
    - Locomotion/Movement
    - Audio Presentation
    - Visual Feedback
    - Visual Surroundings
3. **Which specific features from each of the VR environments did you find particularly engaging?**
4. **Were there elements in either of the environment that you felt were missing or could be improved on?**
5. **How would you compare the two VR environments in terms of ease of use and immersion?**
6. **Can you describe your overall experience using the VR application?**
7. **Which environment did you feel more comfortable in?**
    - Forest Path
    - Space
    - Both
8. **How did you find the experience in making colour and size associations to the auditory stimuli presented?**
9. **What, if any, made you feel comfortable about the environment selected above?**
10. **Did you encounter any technical issues, glitches, or bugs while using the VR application?**
11. **If so, can you describe what the bugs were and in which environment they occurred?**
12. **If you could suggest any changes or additions to enhance your experience in either of the VR environments, what would they be?**

B.6 System Usabilty Scale (SUS)

Instructions: Please indicate how much you agree or disagree with the following statements. Use a 5-point Likert scale:
1 = Strongly Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly Agree

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

B.7 iGroup Presence Questionnaire (IPQ)

Instructions: Please answer the following questions based on your experience in the virtual environment. Use a 7-point Likert scale, with different anchors depending on the question type.
Examples: 1 = Fully Disagree to 7 = Fully Agree; 1 = Not at all to 7 = Very much; 1 = Not real at all to 7 = Completely real

1. In the computer-generated world I had a sense of "being there".
2. Somehow I felt that the virtual world surrounded me.
3. I felt like I was just perceiving pictures.
4. I did not feel present in the virtual space.
5. I had a sense of acting in the virtual space, rather than operating something from outside.
6. I felt present in the virtual space.
7. How aware were you of the real world surrounding you while navigating in the virtual world? (e.g., sounds, room temperature, other people, etc.)
8. I was not aware of my real environment.
9. I still paid attention to the real environment.
10. I was completely captivated by the virtual world.
11. How real did the virtual world seem to you?
12. How much did your experience in the virtual environment seem consistent with your real-world experience?
13. How real did the virtual world seem to you?
14. The virtual world seemed more realistic than the real world.

B.8 Simulator Sickness Questionnaire (SSQ)

Instructions: Please rate the severity of each symptom you experienced after using the VR system. Use a 4-point scale:
0 = None; 1 = Slight; 2 = Moderate; 3 = Severe

1. General discomfort
2. Fatigue
3. Headache
4. Eye-strain
5. Difficulty focusing
6. Increased salivation
7. Sweating
8. Nausea
9. Difficulty concentrating

B.9 Evolution of Template Analysis Themes - Study Two

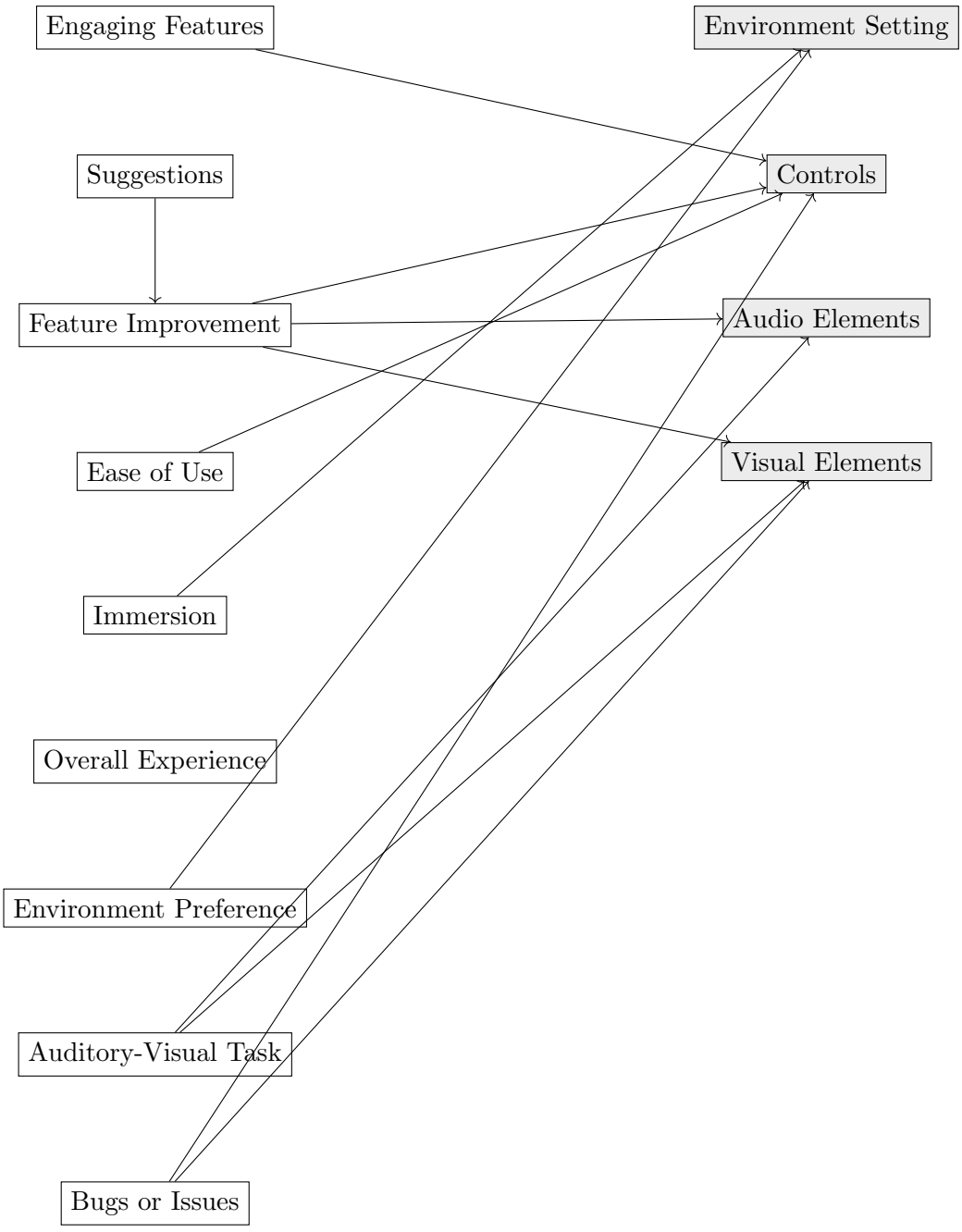


Figure B.2: Visualisation showing how initial *"a priori"* codes (left) contributed to the final template themes (right) during the analysis of Study Two. Some initial codes were linked to multiple final themes where they overlapped in meaning. *"Suggestions"* was combined with *"Feature Improvement"* during coding, and *"Overall Experience"* was excluded from further development.

B.10 Images of the VR Application in Study Two



Figure B.3: An early iteration of the visual feedback in the forest environment in Study Two, using Unity's particle system.

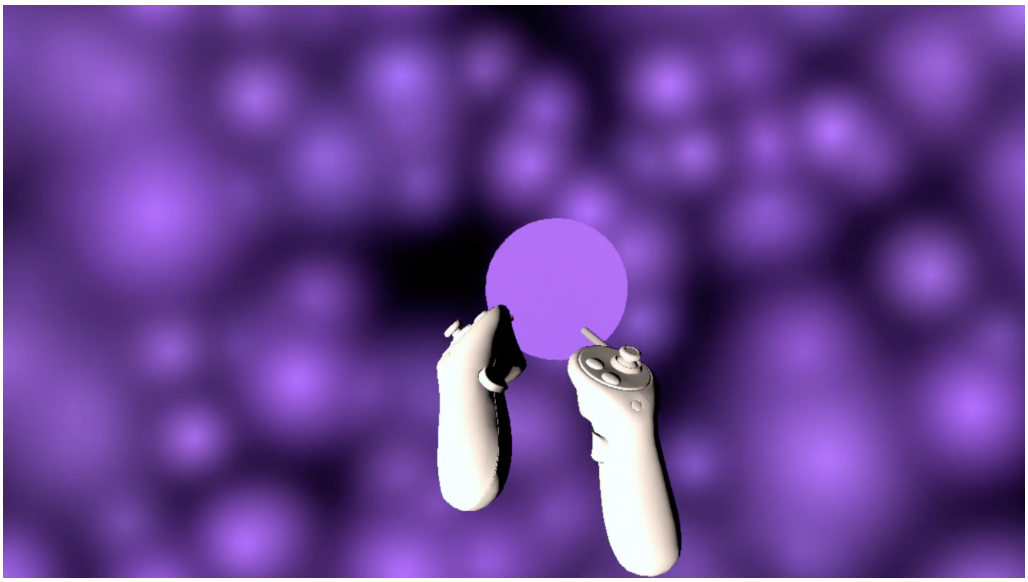


Figure B.4: An early iteration of the visual feedback in the space environment in Study Two, using Unity's particle system.



Figure B.5: First iteration of how to teach the participants the controls needed in the forest environment. The second iteration was to show a video of these interactions at the start of each environment.



Figure B.6: First iteration of how to teach the participants the controls needed in the space environment. The second iteration was to show a video of these interactions at the start of each environment.

Appendix C

Study Three

C.1 Consent Form

Participant no:

- I, ... hereby agree to participate in this usability study to test two VR environments conducted by Patrick O'Toole under the supervision of Dr Ian Pitt and Dr Maurizio Mancini.
- I am 18 years old or older.
- I am participating voluntarily.
- The purpose of the study has been explained to me and I understand what is involved.
- I understand that I can withdraw from the study, without any repercussions.
- I understand that I can withdraw permission to use the data collected from me up to 14 days after my participation in the experiment, in which case the data will be deleted.
- I understand that my information will be kept confidential and my data will be anonymised.
- I give permission for my data to be used in any academic presentation or publication, including the PhD thesis.

Signed: .....
Date: .....

C.2 Information Sheet

This research study consists of learning simple musical theory exercises in an immersive Virtual Reality (VR) environment, which will be presented using a VR head-mounted display. Should you agree to participate in this study, you will be asked to enter the VR environment and complete a short tutorial before starting the exercises.

Control Group Info: [Once you have completed the tutorial, you will start the learning experience. You will listen to musical notes in different sequences and then attempt to replicate the sequence just heard using the auditory-visual interactive objects in front of you.]

Experimental Group Info: [Once you have completed the tutorial, you will start the pre-test environment, where you will need to make sound-colour associations for each object in the environment. When you grab an object, a musical note will play, and you must choose the colour you associate with that sound. When all objects are done, the main exercise will load up, and you will start the learning experience. You will listen to musical notes in different sequences and then attempt to replicate the sequence just heard using the auditory-visual interactive objects in front of you.]

Participation is voluntary, and you can withdraw from the experiment anytime. During the VR task, if you decide to leave the task for any reason, let the researcher know, and they will help you remove the equipment. Afterwards, you will be asked to complete a post-test questionnaire about the VR application and the exercises you participated in.

All information collected from this study will be completely confidential, anonymised and stored on secure University College Cork systems. Each participant will be assigned a participant number, and the questionnaires and sound/colour data will be linked to this number indirectly. There will be no other way to connect the information to the participant. If you decide to withdraw from the research and do not want your data used, you can do so anytime up to 14 days from when the study took place.

The information collected will be used as part of a PhD research project and will be used in academic presentations and publications.

Some potential health issues associated with using head-mounted displays (HMDs) include:

- Eye strain: HMDs can cause eye strain, fatigue, and dry eyes, as the eyes have to focus on a close-up image for an extended period of time.
- Headaches: Some people may experience headaches or migraines due to the weight of the HMD or the strain on the eyes.
- Motion sickness: HMDs can cause motion sickness, as the brain receives conflicting signals from the eyes and inner ear.

If you feel uncomfortable or sick during the Virtual Reality experience as part of this study, please inform the researcher, and they will aid you. Once the researcher is informed, they will help you out of the VR headset, provide you with a comfortable seat and some water, and monitor your condition.

If you feel unwell after you finish the study, here are the contact details for the student health centre.

Tel: +353 (0)21 4902311
Address: Student Health Centre,
Crow's Nest, Victoria Cross Rd.
Cork, T12 HXW4

The study has been reviewed and has received ethical approval from UCC's Social Research Ethics Committee. If there is anything you'd like to know, please contact me or my supervisors directly:

- Patrick O'Toole: patrick.otoole@umail.ucc.ie
- Dr. Laura Maye: laura.maye@ucc.ie
- Dr. Ian Pitt: ianp@cs.ucc.ie
- Maurizio Mancini: m.mancini@di.uniroma1.it

Thank you for your time, please complete the study consent form to confirm your participation in the study.

C.3 Study Three Ethics Approval

Ethics Committee, Social Research <srec@ucc.ie> Tue, 27 Aug 2024, 12:46
to me, Laura

Dear Patrick,

The Social Research Ethics Committee recommends approval of your application, Log No: 2024-159 entitled: **"Understanding Auditory-Visual Associations in Virtual Reality Musical Education Environments."**

- To activate approval, you are required to submit a final revised SREC application form, quoting the log number you received and return a copy to srec@ucc.ie
- The revised application must incorporate all reviewer recommendations (with changes highlighted in yellow text).
- Once the revised form has been resubmitted your approval is active, and you may commence your research.

Please note reviewer one comments:

- Presumably you will be logging identifying participant information with the ID for your own records, so please give appropriate consideration as to how best to ensure data security in this regard. Participants should be assured that their data will be linked to their identity only through an ID code stored confidentially by you.
  - I would suggest storing in your own account until such time as your PhD itself is complete. Once you have finished your PhD your supervisor will become responsible for the storage of your data.

Please note below:

If there are any changes from what has been approved (including but not limited to, end date, list of researchers, recruitment, data collection methods, etc.) you will need to submit an amendment request application to SREC. You can find a copy of the SREC Amendment form on the SREC website under FAQs.

The Committee wishes you the very best with your research.

Kind regards,
Maria

Figure C.1: Approval from the University College Cork Ethics Committee for Study Three presented in Chapter 6.

C.4 Pre-Study Questionnaire Items

1. **What is your Email Address?**
2. **What age group do you belong to?**
    - 18-25
    - 26-35
    - 36-45
    - 46-55
    - 56 or over
3. **What gender do you identify as?**
    - Woman
    - Man
    - Non-binary
    - Prefer not to say
4. **Do you play a musical instrument?**
    - Yes
    - No
5. **If you play a musical instrument, what do you play?**
6. **What is your music level?**
    - Beginner
    - Intermediate
    - Advanced
    - Expert
7. **If yes in Q4, what style of music do you play?**
    - Rock
    - Jazz
    - Country
    - Punk
    - Rap
    - Classical
    - Pop
    - Metal
    - Folk
    - World Music

8. **Do you have any formal music training?**
    - Yes
    - No
9. **If yes to Q7, what kind of training is it?**
10. **Did it include learning musical theory?**
    - Yes
    - No
    - Not sure
11. **Do you have experience using Virtual Reality headsets, like an Oculus Quest or HTC Vive?**
    - Yes
    - No
12. **If yes to Q11, how much would you use a VR headset?**
    - Minimal
    - Occasional
    - Daily
    - Weekly
    - Monthly

C.5 Post-Study Questionnaire Items

1. **User ID:**
2. **Which Group were you assigned to?**
    - Group A
    - Group B
3. **Please give your insights on the pre-test environment where you associated the musical note with colours.**
4. **Did you find it challenging to associate the musical notes with colours? Please explain below.**
5. **Did participating in the pre-test influence your performance in the Music Test? If so, how?**
6. **What was your overall score in the exercise?**
7. **How did you find the difficulty level of the task?**
    - Easy
    - Somewhat Easy
    - Normal
    - Somewhat Difficult
    - Difficult
8. **Based on your assessment in the previous question, please elaborate on both the positive and negative aspects of the Music Test.**
9. **Do you feel you better understand the C Major scale and the intervals within the scale after this exercise? Please explain.**
10. **How did you find using a VR application with a head-mounted display for musical learning purposes?**
11. **Were there any aspects of the VR application that you found particularly compelling or beneficial? Please describe.**
12. **Were there any aspects of the VR application that you found challenging or could be improved? Please describe.**
13. **Do you have any additional insights or suggestions for improving the VR application?**

C.6 Evolution of Template Analysis Themes - Study Three

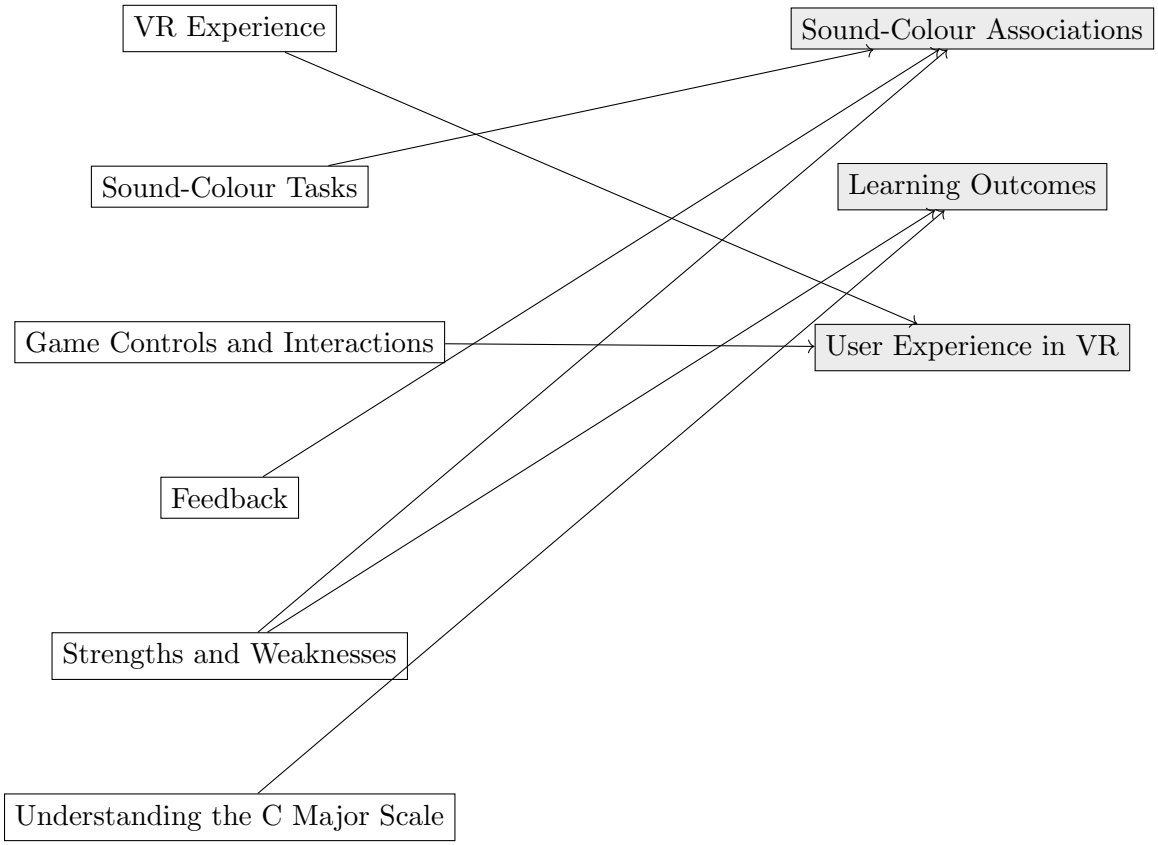


Figure C.2: Visualisation showing how initial *"a priori"* codes (left) contributed to the final template themes (right) during the analysis of Study Three. Several initial codes link to multiple final themes, illustrating how participant feedback informed overlapping aspects of *"Sound-Colour Associations"*, *"Learning Outcomes"*, and *"User Experience in VR"*.

C.7 Images of the VR Application in Study Three

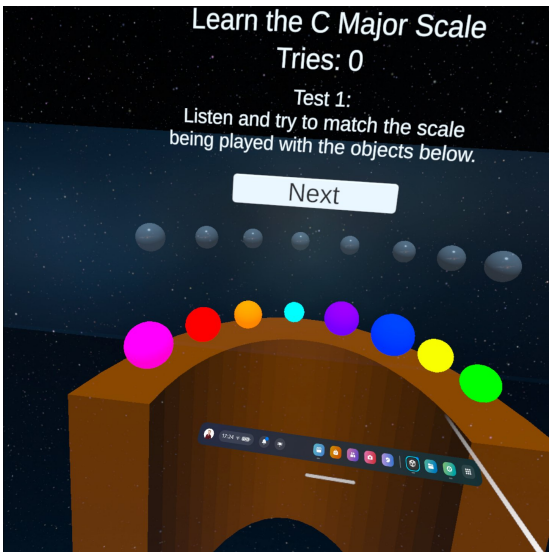


Figure C.3: An early iteration of the music exercise set up in Study Three

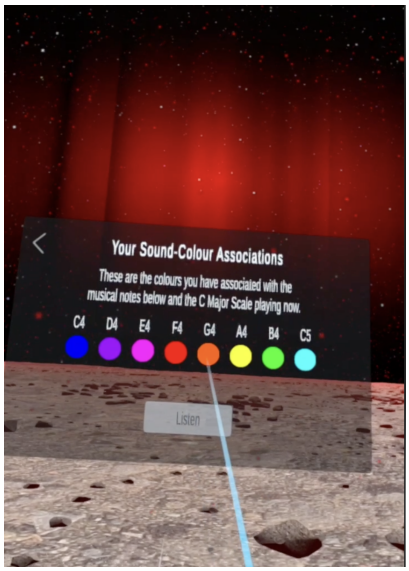


Figure C.4: The reference menu that allows you to check the pre-defined pitch-colour mappings before exercises

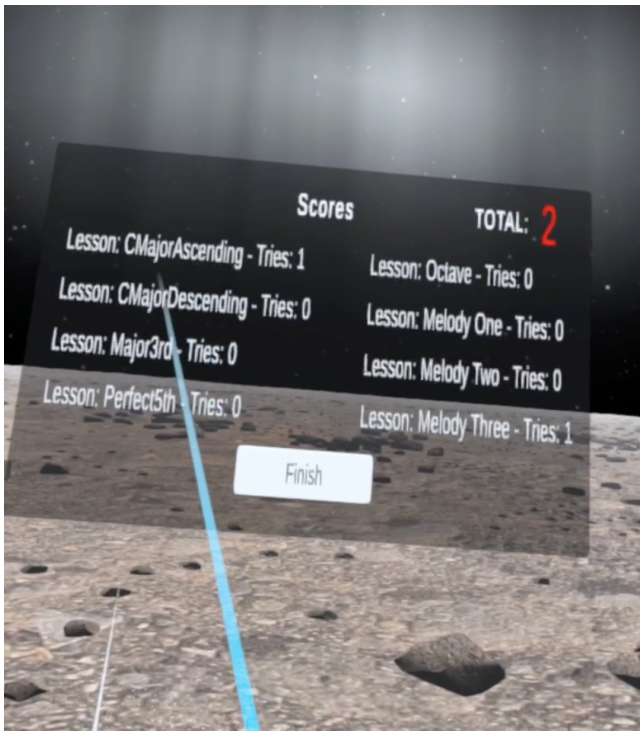


Figure C.5: The menu for the error scores in each lesson and the total that is presented at the end of the exercises.

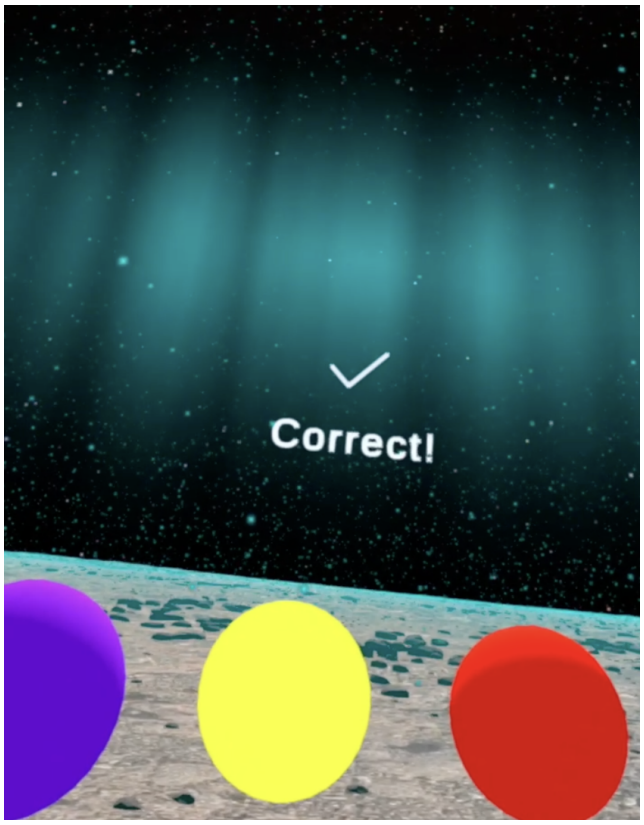


Figure C.6: The Auroa Borealis visual feedback in the music exercises.

C.8 SPSS Output from Study Three

Within-Subjects Effects for Error Scores

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Lesson	Sphericity Assumed	734.000	7	104.857	11.862	<.001	.298
	Greenhouse-Geisser	734.000	3.734	196.556	11.862	<.001	.298
	Huynh-Feldt	734.000	4.534	161.875	11.862	<.001	.298
	Lower-bound	734.000	1.000	734.000	11.862	.002	.298
Lesson * Group	Sphericity Assumed	45.133	7	6.448	.729	.647	.025
	Greenhouse-Geisser	45.133	3.734	12.086	.729	.565	.025
	Huynh-Feldt	45.133	4.534	9.954	.729	.590	.025
	Lower-bound	45.133	1.000	45.133	.729	.400	.025
Error(Lesson)	Sphericity Assumed	1732.617	196	8.840			
	Greenhouse-Geisser	1732.617	104.561	16.570			
	Huynh-Feldt	1732.617	126.962	13.647			
	Lower-bound	1732.617	28.000	61.879			

Figure C.7: SPSS output from the repeated measures ANOVA, looking at the within-subjects effects on the Error Scores.

Within-Subjects Contrasts for Error Scores

Measure: MEASURE_1

Source	Lesson	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Lesson	Linear	107.479	1	107.479	8.541	.007	.234
	Quadratic	170.870	1	170.870	13.215	.001	.321
	Cubic	51.073	1	51.073	5.842	.022	.173
	Order 4	301.896	1	301.896	24.451	<.001	.466
	Order 5	12.862	1	12.862	3.448	.074	.110
	Order 6	24.667	1	24.667	3.223	.083	.103
	Order 7	65.152	1	65.152	16.742	<.001	.374
Lesson * Group	Linear	11.051	1	11.051	.878	.357	.030
	Quadratic	1.829	1	1.829	.141	.710	.005
	Cubic	8.018	1	8.018	.917	.346	.032
	Order 4	4.182	1	4.182	.339	.565	.012
	Order 5	12.198	1	12.198	3.270	.081	.105
	Order 6	6.223	1	6.223	.813	.375	.028
	Order 7	1.633	1	1.633	.420	.522	.015
Error(Lesson)	Linear	352.363	28	12.584			
	Quadratic	362.052	28	12.930			
	Cubic	244.765	28	8.742			
	Order 4	345.717	28	12.347			
	Order 5	104.444	28	3.730			
	Order 6	214.315	28	7.654			
	Order 7	108.961	28	3.891			

Figure C.8: SPSS output from the repeated measures ANOVA, looking at the within-subjects contrasts on the Error Scores.

Between-Subjects Effects for Error Scores

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	2940.000	1	2940.000	28.126	<.001	.501
Group	5.400	1	5.400	.052	.822	.002
Error	2926.850	28	104.530			

Figure C.9: SPSS output from the repeated measures ANOVA, looking at the between-subjects effects on the Error Scores.

Within-Subjects Effects for Time Taken

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Lesson	Sphericity Assumed	177621.850	7	25374.550	28.903	<.001	.508
	Greenhouse-Geisser	177621.850	3.473	51145.384	28.903	<.001	.508
	Huynh-Feldt	177621.850	4.166	42633.369	28.903	<.001	.508
	Lower-bound	177621.850	1.000	177621.850	28.903	<.001	.508
Lesson * Group	Sphericity Assumed	2149.067	7	307.010	.350	.930	.012
	Greenhouse-Geisser	2149.067	3.473	618.814	.350	.818	.012
	Huynh-Feldt	2149.067	4.166	515.826	.350	.851	.012
	Lower-bound	2149.067	1.000	2149.067	.350	.559	.012
Error(Lesson)	Sphericity Assumed	172073.333	196	877.925			
	Greenhouse-Geisser	172073.333	97.241	1769.561			
	Huynh-Feldt	172073.333	116.655	1475.057			
	Lower-bound	172073.333	28.000	6145.476			

Figure C.10: SPSS output from the repeated measures ANOVA, looking at the within-subjects effects on the Time Taken.

Within-Subjects Contrasts for Time Taken

Measure: MEASURE_1

Source	Lesson	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Lesson	Linear	16228.956	1	16228.956	10.337	.003	.270
	Quadratic	48931.670	1	48931.670	94.698	<.001	.772
	Cubic	16890.449	1	16890.449	25.985	<.001	.481
	Order 4	58039.096	1	58039.096	64.579	<.001	.698
	Order 5	9865.381	1	9865.381	14.240	<.001	.337
	Order 6	4096.517	1	4096.517	7.208	.012	.205
	Order 7	23569.781	1	23569.781	18.874	<.001	.403
Lesson * Group	Linear	25.429	1	25.429	.016	.900	.001
	Quadratic	1018.801	1	1018.801	1.972	.171	.066
	Cubic	241.851	1	241.851	.372	.547	.013
	Order 4	309.614	1	309.614	.344	.562	.012
	Order 5	419.392	1	419.392	.605	.443	.021
	Order 6	113.952	1	113.952	.200	.658	.007
	Order 7	20.028	1	20.028	.016	.900	.001
Error(Lesson)	Linear	43961.103	28	1570.039			
	Quadratic	14467.922	28	516.712			
	Cubic	18199.966	28	649.999			
	Order 4	25164.579	28	898.735			
	Order 5	19398.841	28	692.816			
	Order 6	15914.099	28	568.361			
	Order 7	34966.823	28	1248.815			

Figure C.11: SPSS output from the repeated measures ANOVA, looking at the within-subjects contrasts on the Time Taken.

Between-Subjects Effects for Time Taken

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	882578.817	1	882578.817	215.457	<.001	.885
Group	38.400	1	38.400	.009	.924	.000
Error	114696.533	28	4096.305			

Figure C.12: SPSS output from the repeated measures ANOVA, looking at the between-subjects effects on the Time Taken.

Appendix D

Miscellaneous

D.1 Colour Comparison Between a VR HMD and Desktop

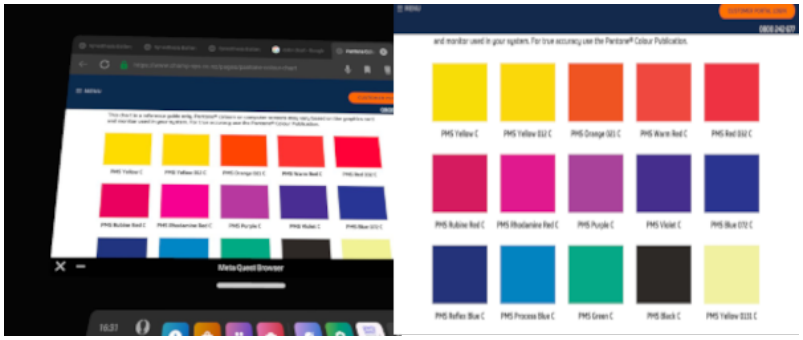


Figure D.1: A manual visual test of how the colours were presented in VR (on the left) and Desktop (on the right) set-up was conducted in the design stage for Study One presented in Chapter 4.

D.2 Test of an Alternative Browser in VR

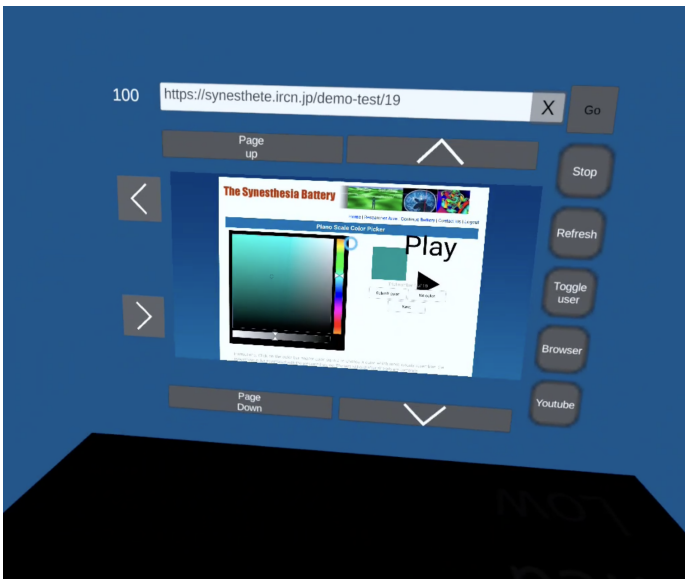


Figure D.2: This image shows a test of an alternative VR browser (`UnityAndroidVRBrowser`[a]) evaluated for potential use in Study One (Chapter 4). Although it was not used in the final implementation, it was explored as part of the early testing process on the `Oculus Quest 2`.

[a] UnityAndroidVRBrowser, `https://github.com/IanPhilips/UnityOculusAndroidVRBrowser`, accessed 19 January 2023

D.3 VR HMD Specs for the Quest 2 and Newer

Table D.1: Comparison of VR Headsets Released Between 2020-2024

Headset	Release Date	Display (per eye)	Processor	Notable Features	Price Range
Meta Quest 2	Oct 2020	1832×1920 LCD	Snapdragon XR2 Gen 1	Standalone VR, 6DoF tracking, hand tracking	$299-$399
Meta Quest 3	Oct 2023	2064×2208 LCD	Snapdragon XR2 Gen 2	Full-colour MR, improved optics	$499-$649
Meta Quest 3S	Oct 2024	1832×1920 LCD	Snapdragon XR2 Gen 2	Affordable MR, upgraded processor	$299-$399
Meta Quest Pro	Oct 2022	1800×1920 MiniLED LCD	Snapdragon XR2+ Gen 1	Eye/face tracking, enterprise focus	$1,499
PICO 4	Oct 2022	2160×2160 LCD	Snapdragon XR2	Lightweight design, high-resolution display	€429-€499
HTC Vive XR Elite	Feb 2023	1920×1920 LCD	Snapdragon XR2	Modular design, full-colour passthrough	$1,099
PlayStation VR2	Feb 2023	2000×2040 OLED	Custom (PS5 integration)	Eye tracking, haptic feedback	$549.99
Apple Vision Pro	2024	3840×3744 Micro-OLED	Apple M-series	Advanced tracking, spatial audio	$3,499

Appendix E

Publications and Contributions

The following publications have resulted from the research conducted as part of my PhD. Three of these papers (see P5, P7, and P8) form the main body of this thesis by publication.

P1 **P. O'Toole**, D. Glowinski, I. Pitt, M. Mancini. *When Emotions are Triggered by Single Musical Notes: Revealing the Underlying Factors of Auditory-Emotion Associations.* International Conference on Multimodal Interaction (ICMI), 2021, pp. 291-298. DOI: 10.1145/3461615.3485419.

My Contribution: As lead author, I designed the questionnaires, recruited participants, and assisted in the quantitative and qualitative analysis on the results. I wrote the first draft of the paper, and contributed to subsequent edits. My co-authors were closely involved at all stages.

P2 B. Biancardi, **P. O'Toole**, I. Giaccaglia, B. Ravenet, I. Pitt, M. Mancini, and G. Varni. *How ECA vs Human Leaders Affect the Perception of Transactive Memory System (TMS) in a Team.* 9th International Conference on Affective Computing and Intelligent Interaction (ACII), 2021, pp. 1-8. DOI: DOI: 10.1109/ACII52823.2021.9597454.

My Contribution: As co-author, I designed the survey, and contributed to subsequent edits of the paper.

P3 **P. O'Toole**. *Understanding Personalised Auditory-Visual Associations in Multi-Modal Interactions.* International Conference on Multimodal Interaction (ICMI), 2021, pp.812-816. DOI: 10.1145/3462244.3481277.

My Contribution: As lead author, I wrote the first draft of the paper, and contributed to subsequent edits.

P4 **P. O'Toole**. *Can VR Help Understand Auditory-Visual Associations and Synaesthesia Through Immersive Battery Tests?.* 10th International Conference on Affective Computing and Intelligent Interaction (ACII), 2022, pp. 1-5. DOI: 10.1109/ACIIW57231.2022.10086003.

My Contribution: As lead author, I wrote the first draft of the paper, and contributed to subsequent edits.

P5 **P. O'Toole**, M. Mancini, and I. Pitt. *Sound and Colour: Evaluating Auditory-Visual Tests in Virtual Reality and Traditional Desktop Settings.* The International Conference on Advanced Visual Interfaces (AVI), 2024, pp. 1-5. DOI: 10.1145/3656650.3656665.

My Contribution: As lead author, I designed the questionnaires, recruited participants, and undertook quantitative and qualitative analysis on the results. I wrote the first draft of the paper, and contributed to subsequent edits. My co-authors were closely involved at all stages.

P6 **P. O'Toole**, M. Mancini, I. Pitt, L. Maye. *Chromaesthetica: A Virtual Reality App for Learning Musical Theory Using Sound-Colour Cross-Modal Associations.* The International Conference on Mobile and Ubiquitous Multimedia (MUM), 2024, pp. 1-3. DOI: 10.1145/3701571.3703369.

My Contribution: As lead author, I designed the VR demonstration, wrote the first draft, and contributed to edits.

P7 **P. O'Toole**, M. Mancini, I. Pitt, and L. Maye. *Investigating Musical Pitch-Colour Associations in Immersive Virtual Reality Environments.* Virtual Reality Springer Journal 29, 88, 2025, pp. 1-23. DOI: 10.1007/s10055-025-01163-8.

My Contribution: As lead author, I designed the VR application and the questionnaires, recruited participants, and undertook quantitative and qualitative analysis on the results. I wrote the first draft of the paper, and contributed to subsequent edits. My co-authors were closely involved at all stages.

P8 **P. O'Toole**, M. Mancini, I. Pitt, and L. Maye. *Exploring Cross-Modal Sound-Colour Associations for Music Learning in Immersive Virtual Reality.* Submitted to the *Computers & Education: X Reality Journal.*

My Contribution: As lead author, I designed the VR application and the questionnaires, recruited participants, and undertook quantitative and qualitative analysis on the results. I wrote the first draft of the paper, and contributed to subsequent edits. My co-authors were closely involved at all stages.