Developing a Musical Practice in XR: Music Composition, Performance and Improvisation Along the Virtual Continuum

Matias Vilaplana Stark
Santiago, Chile

M.A., University of Virginia, 2023
M.A., University of Michigan, 2019
B.A., Universidad de Chile, 2016

A Dissertation presented to the Graduate Faculty
of the University of Virginia for the Degree of Doctor of Philosophy

Department of Music

University of Virginia
May 2025

Luke Dahl (chair), Ph.D.
Ted Coffey, Ph.D.
JoVia Armstrong, Ph.D.
Mona Kasra, Ph.D.

Abstract

Extended reality (XR) technology has become widely available to consumers in recent years, with ‘tech’ companies releasing new headsets and controllers for virtual and augmented reality on almost a yearly basis. The wider access to this technology presents a myriad of opportunities for artists and researchers to develop novel musical instruments and musical experiences using 3D interactive immersive environments. Various design frameworks have been proposed for the creation of musical instruments and experiences with XR technology, however, this is still a developing field when considering the output of creative works that take an audio-first approach or focus on music composition, improvisation and performance more specifically. What can XR contribute to the development of existing musical practices? What new musical practices can be developed through XR? How can these technologies change our perception, expectations and shape our mental models regarding making and experiencing music? This dissertation consists of a written document and a series of XR musical pieces that were developed over the course of the Ph.D. Each of these pieces has served as an exploration of the different possibilities and creative routes afforded by XR technology, and the challenges to the design, composition and performance of musical pieces that range from fully immersed individual experiences to ensemble performances that incorporate virtual environments.

Acknowledgements

If I could, I would list everyone I have crossed paths with in my life to thank them. Because every interaction, every conversation, every shared moment, has shaped my journey in one way or another.

I want to thank my family for their unconditional love and support. My mother, Andrea, for always encouraging me to pursue my dreams and to overcome my limitations—in the same ways that you do. My father, Alberto, for supporting me and ultimately believing in my life choices. My brother, Benjamin, for rooting for me and for sharing your artistic sensitivity. And my grandfather Jorge, for teaching me the true value of life and for being my biggest inspiration. You taught me to move forward without fear, and I know you are always watching over me.

I am very grateful for all of the professors and mentors that I have met during different stages of my academic journey. First, I want to thank Luke for his guidance as my advisor all these years at UVA. I will miss our conversations that always seem to go in multiple directions, from music to technology, to theory, and the body, to music gear or an album. Thank you for inviting me to participate in many exciting and challenging projects these past years. Ted, many things you have said to me over the years–whether in class or in passing–have stayed with me and your insights on music (and beyond) continue to resonate with me. I owe a great debt having received your wisdom. Thank you, JoVia, for lending an ear to my frustrations and sharing your experiences to guide me through the process of writing this dissertation. Nicole, you have opened new possibilities for me in writing music. Your class and your mentorship have pushed me outside of my comfort zone in performance, and shown me new ways to listen when playing music with others. Noel, thank you for building community everywhere you go, and for showing all of us the importance of that work in academia and in life. I am grateful for the opportunity you have given me to teach with you, to learn from the artists at the Black Power Station, and for all the long conversations on music, culture, and of course, football (soccer for my US friends). Anıl, thank you for introducing me to immersive media, for pushing me further in our collaborations, and for helping me navigate the academic opportunities in the world of music and XR. Michael, thank you for helping me take my first steps into the world of interactive design as my advisor at Michigan, and for our collaborations while at UVA. Thank you, John, for inviting me to collaborate in different performance and research projects (and for letting me stay at 'the shack' in the early pandemic months). Javier, I am grateful for your guidance during my last years as an undergraduate. Thank you for showing me how much broader the world of music technology is, for trusting me as a collaborator and for encouraging me to pursue my creativity and academia. I hope to cross paths again with all of you, whether in music, scholarship, or simply as people, I will always be grateful for the impact you all had on me.

I want to thank all of my peers in the music department with whom I have had the opportunity to share gatherings, meals (especially fondue), concerts, ensembles, performances, colloquiums, and conferences. Thank you Qi, Neda, Gabrielle, Rah, Molly, Kristin, Varun, Brian, Rachel, Corey, Sia, Emily, Hannah, Sam, Daniel, Juan, and Omar. Thank you Alex for always keeping your office door open, for always being supportive, for all the great discussions on art and music, for always lending thoughtful feedback on my work, and for helping me navigate these last years of the Ph.D while going to get coffee or bagels. A special shoutout to my housemates at Forest Hills: Basile, Dilshan, Becky, and Heather. You all made Charlottesville a better place during the early pandemic years, thank you for all the laughs and fun moments, for all the times we celebrated and supported each other, for all the dinners, barbecues, movie nights, and jam sessions at that beloved place we called home for a few years.

Thank you to all the good friends I have made throughout these years both in Chile and the US. Thank you Simón, for your warmth, your joy, and your excitement for music making. Thank you Santiago, for all the work we did together, for always connecting musicians with each other, and for the pursuit of meaningful art. Thank you Esteban, for always lifting your classmates and for your exemplary work ethic. To my fellow bandmates, Felipe and Diego, thank you for introducing me into music making, and for being my second family down the street for so many years. To Juan Felipe, for being a good friend and helping me navigate moving to the US. Thank you to my friends from Ann Arbor: Morgan, Joey, Kate, Jake, Jacob, and Ryan. Thank you for welcoming me into the US, for all the music we made, the laughter, the good times, and all the adventures. To my friend Anish, thank you for joining me at ‘the shack’ in those early pandemic months that were filled with uncertainty and quiet chaos, although it is a blurry time, I will never forget all the great meals we made and the movies we watched. To my friends in Charlottesville, Jesús, Pato, Javier, Lucas, and Adrian; thank you for providing a space where we can speak and joke around in our native tongues, oftentimes with football in the background or in an outdoors bar.

Last but not least, I would like to thank my wife Tracey. Thank you for your unconditional love throughout these years in graduate school. Thank you for showing me the little things that matter, for letting me into your life, for opening my heart to the simple things, and for pushing me to be more rigorous in my work. I would not be who I am today without your thoughtfulness, your companionship, and your support throughout this whole process.

Table of contents

Introduction **1**

Chapter 1
A historical overview of XR and computer music **3**

- **1.1 The Visual and Sonic modalities in XR** **4**
- **1.2 XR technology** **6**
  - 1.2.1 Military research 6
  - 1.2.2 Creative research 8
  - 1.2.3 Entertainment industry 10
  - 1.2.4 Current technologies 11
- **1.3 Sonic Simulations: Disembodied Sound** **13**
  - 1.3.1 Musique Concrète and Elektronische Musik 14
  - 1.3.2 Digital Sound Synthesis 14
  - 1.3.3 Physical Modeling 18
  - 1.3.4 Granular synthesis: simulating environmental textures 19
- **1.4. Sonic Simulations: Spatial Audio, simulating space and movement** **21**
  - 1.4.1 Early experiments in recording and playback 21
  - 1.4.2 Multi-channel spaces and spatial audio formats 22
- **1.5 Sonic Simulations: Musical Agents** **24**
  - 1.5.1 Networks and computer players 24
  - 1.5.2 George E. Lewis and Voyager 25
- **1.6 Sonic Simulations: Artificial Intelligence** **28**
  - 1.6.1 Generative models and the problem of digital representation in music 28
  - 1.6.2 AI music generation and Neural Audio Synthesis 29
- **Chapter 1 Summary** **31**

Chapter 2
The XR musical landscape: Concepts, Design Frameworks, and Creativity **32**

- **2.1 XR Concepts** **32**
  - 2.1.1 Immersion 32
  - 2.2.2 Presence 33
  - 2.2.3 Virtual Embodiment 35
  - 2.2.4 Cybersickness 37
- **2.2 Design and evaluation frameworks in musical XR** **38**
  - 2.2.1 Designing musical instruments for virtual environments 38
  - 2.2.2 Aesthetic driven design 38
  - 2.2.3 Scenography for XR 39
- **2.3 Review of creative works in XR** **39**

2.3.1 Traditional Musical Instruments.......... 41
2.3.1.1 Virtual Percussion.......... 42
2.3.1.2 Virtual String Instruments.......... 42
2.3.2 Virtual Navigation.......... 43
2.3.2.1 Floating in a virtual environment.......... 44
2.3.2.2 Directed navigation as performance.......... 45
2.3.2.3 ‘Open world’ navigation as a musical instrument.......... 45
2.3.2.4 3D space as a sequencer.......... 46
2.3.3 3D Interactions and Controllers.......... 46
2.3.3.1 Custom physical controllers for interactions in immersive environments.......... 47
2.3.3.2 Technologies for immersive performances.......... 47
2.3.3.3 Re-arranging scores in VR.......... 48
2.3.3.4 Magnet based haptics for VR music controllers.......... 48
2.3.4 Shared Virtual Environments.......... 49
2.3.4.1 Gesture control in multiplayer music environments.......... 49
2.3.4.2 Modular synthesis in VR.......... 50
2.3.4.3 Exploring communication for music performance in shared virtual environments.......... 50
2.3.4.4 Performing in virtual spaces.......... 51
2.3.5 Gamification and Video Game Mechanics.......... 52
2.3.5.1 Repurposing video game mechanics for music composition.......... 52
2.3.5.2 Game aesthetics in virtual instrument design.......... 52
2.3.5.3 Custom game mechanics for music composition.......... 53
2.3.5.4 Video games as scores.......... 53
2.3.6 Sandbox environments and audiovisual programming.......... 54
2.3.6.1 Immersive sandboxes.......... 54
2.3.6.2 Audiovisual programming.......... 55
2.3.6.3 Multisensory programming.......... 55
2.3.7 Live performances in XR.......... 56
2.3.7.1 Multi-user mixed reality.......... 56
2.3.7.2 VR conductor.......... 57
2.3.7.3 Augmenting reality through sound.......... 57
2.3.7.4 Sampling sound, indexing space.......... 58
2.4 Emerging Trends in Composition in Musical XR.......... 58
2.4.1 Musical form.......... 59
2.4.2 Performing music through spatial interactions in XR.......... 62
2.4.3 Virtual re-embodiment of sound.......... 65
2.4.4 Experiencing XR music compositions.......... 68
Chapter 2 Summary.......... 70

Chapter 3
Music composition and performance along the virtuality continuum.. 71
3.1 VR-Mapper: a Max For Live device for prototyping movement-based sound interactions in virtual reality..71
3.1.1 Touchless interfaces and movement-based interactions.. 71
3.1.2 Implementation.. 73
3.1.2.1 Hardware and software platforms..73
3.1.2.2 Movement tracking and controller input... 73
3.1.2.3 Developing a Max for Live device.. 74
3.1.3 *UX-1*, experimental video performance.. 76
3.1.3.1 Recording movement and sound.. 77
3.1.3.2 Performing a broadcast software.. 78
3.2 Push & Pull and pedalGround.. 80
3.2.1 Design framework, embodied interaction design for virtual environments..........................80
3.2.2 Implementation.. 82
3.2.2.1 Hardware and software platforms..82
3.2.2.2 System mechanics and controller mapping... 83
3.2.3 *Push & Pull*...85
3.2.3.1 Composing space...87
3.2.4 *pedalGround*..88
3.2.4.1 Mixed-reality interactions...89
3.2.4.2 Gamified music improvisation... 91
3.3 *Unforeseen Collisions*: music collaboration in shared virtual environments..........................93
3.3.1 Design Framework...94
3.3.2 Virtual environment and system mechanics...95
3.3.3 Hardware and Software Implementation... 97
3.3.4 Network Design.. 97
3.3.5 Live music performance with *Unforeseen Collisions*..98
3.3.5.1 Online Performance...98
3.3.5.2 Hybrid Performance.. 99
3.3.6 Musical improvisation and composition in *Unforeseen Collisions*.................................... 100
3.4 Virtual environments as graphic scores for music improvsiation..102
3.4.2 Senderos: A dynamic 3D graphic score..103
3.4.2.1 Implementation..104
3.4.2.2 Mapping terrain features to musical parameters..105
3.4.3 Paisajes Oníricos/Dreamscapes: interactive 3D graphic score... 106
3.4.3.1 Visual design and 3D modeling...107
3.4.3.2 Implementation.. 111
3.4.3.3 Virtual Conductor.. 112
3.5 Immersive storytelling: multi-channel soundscape composition...115
3.5.1 Animal grief, soundscape composition, and participatory listening.................................. 115

3.5.3 *Funeral For A Whale*: Structure and Description ... 116
3.5.3.1 Section I, Underwater ... 118
3.5.3.2 Section II, Decay & Corpse Surfacing ... 119
3.5.3.3 Section III, Procession and Ascension ... 120
3.5.4 Spatial audio tools, sound materials and audio processing ... 121
3.5.4.1 Time stretching/compression and transposition ... 121
3.5.4.2 Cleaning sound samples ... 122
3.5.4.3 Layering and spatialization ... 122
3.6 3D Corpus Navigator: a virtual musical instrument for sound corpus manipulation ... 124
3.6.1 Interactive systems that use Corpus Based Concatenative Synthesis ... 125
3.6.2 Implementation ... 126
3.6.3 Interface for descriptor analysis and general settings ... 126
3.6.3.1 Segmentation, Descriptor Analysis, 2D and 3D plotting ... 127
3.6.3.2 Movement and playback settings ... 128
3.6.4 Virtual environment and controls ... 128
3.6.4.1 Movement and interaction controls ... 129
3.6.4.2 Sonic interaction ... 130
3.6.5 Composing and Performing with the 3D Corpus Navigator ... 131
3.6.5.1 Curating a sound corpus ... 131
3.6.5.2 Sonic extension, audio effects and mapping ... 132
Chapter 3 Summary ... 136
Conclusion ... 137
Appendix ... 141
Links to compositions, performances and demos ... 141
Scores ... 142
Senderos ... 142
Paisajes Oníricos / Dreamscapes ... 146
Bibliography ... 154

List of Figures and Tables

Figure 1. Virtuality Continuum as proposed by Milgram & Kishino……………………………………5
Figure 2. Relationships between categories in XR musical practice…………………………………...41
Figure 3. VR-Mapper, Max for Live device…………………………………………………………75
Figure 4. VR-Mapper, one-to-one and many-to-one mappings……………………………………...76
Figure 5. Still frame used in *UX-1* from saturated computer desktop recording…………………………78
Figure 6. Video recordings organized in OBS………………………………………………………78
Figure 7. Instances of overlaying audiovisual recordings performing with VR-Mapper………………....79
Figure 8. Data structure for *Push & Pull* (a) and *pedalGround* (b)………………………………………82
Figure 9a. System physics mechanics……………………………………………………………..84
Figure 9b. Controller input………………………………………………………………………...84
Figure 10a. Landscape in *Push & Pull*………………………………………………………………85
Figure 10b. Top view of spatial arrangement of virtual objects……………………………………...85
Figure 11a Target virtual objects in *Push & Pull*……………………………………………………....85
Figure 11b Different sized spheres controlled by the user…………………………………………..85
Figure 12. View of the virtual environment in *pedalGround*…………………………………………...88
Figure 13a. Target virtual objects in *pedalGround*…………………………………………………..89
Figure 13b. Different sized spheres controlled by the user……………………………………………..89
Figure 14. System configuration for mixed-reality interactions in *pedalGround*…………………………90
Figure 15. Virtual camera pointing towards the two performers in the *Unforeseen Collisions* play area……………………………………………………………………………………………..95
Figure 16. Resonant blocks on a pedestal in *Unforeseen Collisions*…………………………………...96
Figure 17. Hybrid performance. In-person performer (left). Remote performer (right)…………………99
Figure 18. Technical setup for hybrid performance……………………………………………………..100
Figure 19. Satellite images used for *Senderos*………………………………………………………...103
Figure 20. Satellite images rendered as 3D landscapes in a virtual environment……………………….104
Figure 21. Engravings from *Perspective* by Hans Vredeman De Vries…………………………………107
Figure 22. Top view of the virtual environment in Blender……………………………………………...108
Figure 23. Side view of the virtual environment in Blender……………………………………………..108
Figure 24. Snapshots from the conductor's first person perspective of the long narrow hall……………109
Figure 25. Top view of section 2 (left). 3D objects in the ground floor corridors (right)………………..109
Figure 26. Models in Tinkercad (left). Top view of the third section's maze (right)……………………110
Figure 27. Conductor's first person view from inside of the third section maze………………………...110
Figure 28. Setups for live performance: (1) computer running the virtual environment, (2) screen for the audience, (3) screen for performers on stage………………………………………………………112
Figure 29. Mapping of music parameters to the red beams in movement #1……………………………113
Figure 30. Mapping of music parameters to the vortex 3D objects in movement #2……………………114
Figure 31. Conductor's first person view of the maze for movement #3……………………………..........114
Figure 32. Spectrogram of *Funeral For A Whale*……………………………………………………..117
Figure 33. Spectrogram of section I of *Funeral For A Whale*………………………………………….118
Figure 34. Spectrogram of section II of *Funeral For A Whale*…………………………………………..119

Figure 35. Spectrogram of section III of *Funeral For A Whale*…………………………………………120
Figure 36. 3D Corpus Navigator Max/MSP interface © 2025 IEEE…………………………………….127
Figure 37. 3D point cloud in the 3D Corpus Navigator virtual environment……………………………129
Figure 38. Game-pad controller input mapping for the 3D Corpus Navigator © 2025 IEEE…………...130
Figure 39. Mapping #1, stereo delay + bandpass filter…………………………………………………..133
Figure 40. Mapping #2, comb filters + ring modulators…………………………………………………134
Figure 41. Mapping #3, harmonizer + stereo delay……………………………………………………...135
Figure 42. Mapping #4, harmonizer + comb filter……………………………………………………….135

Table 1. Data from hand-held controller input……………………………………………………………74
Table 2. Virtual object and virtual instrument mapping…………………………………………………..86
Table 3. Pitch mapping between spheres and virtual instruments……………………………………...87
Table 4. Target to audio effect mapping…………………………………………………………………..90
Table 5. Summary of musical parameters associated with each landscape in *Senderos*…………………105

Introduction

Composer Marko Ciciliani writes about expanding the practice of music composition through the inclusion of 'non-sonic elements' to express musical ideas where 'sound alone is no longer sufficient.'[1] In his writing he discusses the role of musical instruments, and how the creation of new musical instruments can be considered part of music composition, while identifying a shift towards 'media-oriented' instruments and the influence of the 'logic of the computer' in their development.[2] He argues that composers working in the 'expanded field' have to learn different sets of skills that depart from the music education in Western musical tradition. Which skills a composer will develop in their expanded field will depend on the specific media that composer chooses to engage in. This can range from programming to photography, staging, lighting, soldering, among many others. Ciciliani argues that learning these skills has an impact on the understanding of the field of music itself, moving away from the homogenous discourse of *New Music*, and presenting itself in what Joana Demers refers to as *discursive accents*,[3] where multiple influences and references from different fields are combined.

It is within this framework that my musical practice using eXtended Reality (XR) technology is inscribed into. In developing my work, I have acquired a variety of skills including: programming, 3D modeling, graphic design, video game design, and spatial audio. These skills have allowed me to design and implement interactive systems to perform the different creative works that are discussed in this dissertation. My contribution with this document is to present my creative practice in musical XR in the light of existing design frameworks, creative works by other artists and researchers, and identify compositional trends in this medium in order to continue developing works in XR as an expanded field of music. The document is divided into three chapters that discuss the historical precedents in the development of this technology, the existing creative work in music in XR, and my own musical work with these technologies.

[1] Ciciliani, Marko. "Music in the expanded field: On recent approaches to interdisciplinary composition." *Darmstädter Beiträge zur Neuen Musik* 24 (2017): 23-35.
[2] Manovich, Lev. *The language of new media*. MIT press, 2002.
[3] Demers, Joanna. "Discursive Accents in Some Recent Digital Media Works." (2013).

Chapter one draws a connection between the historical developments of XR display technology and sound synthesis in computer music by identifying common themes relating to virtuality and simulation. While developments in XR were for the most part concerned with the visual display, of 3D graphics, and visual immersion; computer music was concerned with the development of synthesis algorithms that could mimic the timbre of real-world instruments. Although the technologies involved have different goals and challenges for their respective fields, ultimately both areas have developed computational systems to create more accurate virtual representations of visual and acoustic phenomena in the real world.

In chapter two the convergence of XR and computer music is presented through an overview of relevant concepts, design frameworks, and creative works in musical XR. The chapter organizes and describes the different works of artists and scholars through different themes, including: 3D interactions and controllers, virtual navigation, shared virtual environments, gamification and video game mechanics, sandbox environments and audiovisual programming, and live performances in XR. The chapter discusses the different compositional trends identified across the creative works in the different themes and how they relate to ideas of musical form and expression.

Chapter three describes and discusses multiple creative works developed as part of this dissertation. This includes the use of immersive technologies for the development of interactive music systems for music composition and performance in XR, virtual environments as dynamic 3D graphic scores, collaborative virtual environments for music making, and multi-channel soundscape composition.

Chapter 1
A historical overview of XR and computer music

Within the last decade we have witnessed the expansion of eXtended Reality (XR) technology through the commercialization of a wide variety of products available to the public. This has opened possibilities for new artistic practices with immersive technologies that extend beyond historical applications of this technology in the military and entertainment industries. Nowadays, we frequently encounter images of people interacting with these technologies in both private and public spaces. It is not uncommon to see a person wearing a headset protruding from their face while moving their arms to perform rigid and direct gestures in silence. Or to see someone pinching their thumb and index finger into the thin air while wearing a headset to interact with a virtual space that we cannot see. Some applications allow us to point the camera on our mobile device towards the living room and overlay a life-size 3D model of a new couch to see if it would fit before ordering it. Museums and galleries have implemented QR codes that attendees can scan to hear an auditory description of an art piece in a museum. All of these examples reveal interactions that link the real world with virtual elements through technologies that entangle visual representations, gestural interactions, and auditory cues.

All of the examples I have described can be classified under the term of XR.[4] An umbrella term usually employed to refer to virtual reality (VR), augmented reality (AR), and mixed reality (MR). Researchers have contested the term, in an attempt to think more broadly about how our reality is "extended" or "crossed"[5] by technology to include physical devices (glasses, sleep masks), mobile computing, sensing devices, intersecting on topics of surveillance, Mediated Reality, Artificial Intelligence (AI) and Humanistic Intelligence (HI).[6]

---

[4] Stanney, Kay M., et al. "Extended reality (XR) environments." *Handbook of human factors and ergonomics* (2021): 782-815.
[5] Paradiso, Joseph A., and James A. Landay. "Guest editors' introduction: Cross-reality environments." *IEEE Pervasive Computing* 8.3 (2009): 14-15.
[6] Mann, Steve, et al. "All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality." *arXiv preprint arXiv:1804.08386* (2018).

1.1 The Visual and Sonic modalities in XR

Throughout our recent history, the development of XR technology has been largely focused on the visual modality, with auditory perception relegated to a secondary category that mostly supports and reinforces the visual. This is clearly exemplified in the 'virtuality continuum' (Figure 1.), a taxonomy of visual displays for immersive technologies by researchers Milgram & Kishino.[7] Visual displays are sorted along this continuum with the real world on the left end, and a virtual environment on the right end. The latter can be understood as a completely synthetic space with virtual objects and virtual representations of ourselves. Everything in between the two ends of the continuum falls under the category of "mixed reality" where both real and virtual elements interact.

Virtual reality (VR) systems would fit in the right end of the continuum. These systems often use a Head Mounted Display (HMD) with some form of head-tracking to enable users to have a 360º view of virtual space. In VR, users will most commonly navigate virtual space using hand-held controllers. The HMD in this case occludes any visual stimulus from the outside world.

Augmented reality (AR) systems on the other hand, display elements of both the real and virtual world by overlaying virtual objects in physical space. This is achieved through cameras in mobile devices (smartphones and tablets), wearables (smart glasses), or through dedicated HMDs with passthrough views (via cameras) of the real world, allowing the superimposition of virtual objects into the real world.

Augmented virtuality (AV) is a less common and rare case, but one could imagine an immersive system with computer generated graphics that contains real world elements in it, for example, holding a real cup of coffee while immersed in a virtual spaceship.

Mixed reality (MR), is considered by the authors as a display paradigm in which real and virtual objects are presented together simultaneously, which is why AR and AV fall under this category.

[7] Milgram, Paul, and Fumio Kishino. "A taxonomy of mixed reality visual displays." *IEICE TRANSACTIONS on Information and Systems* 77.12 (1994): 1321-1329.

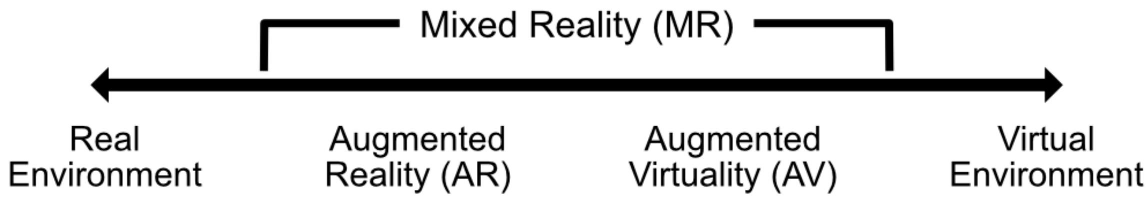


Figure 1. Virtuality Continuum as proposed by Milgram & Kishino

However, as mentioned earlier, this framing of XR and the distinctions between VR, AR and MR displays, relies heavily on the visual modality and falls short in acknowledging the impact tactile and auditory modalities have in XR experiences. Researchers and artists have argued for music-driven discussion and musician-led creative research to deliver novel, enhanced forms of musical expression with an audio-first approach to VR.[8] Although the role of sound in enhancing the sense of presence in virtual environments is well acknowledged, its role needs to be expanded from complimentary to being given the same relevance as the visual components in the development of XR experiences. For example, this approach would provide a better understanding to develop VR experiences for the visually impaired. In the emerging field of Musical XR, researchers have expressed that:

> "The roles of sound and music within an XR experience can range widely from passive background elements to fully interactive and controllable phenomena, where sound acts as a fundamental driver of, is driven by, or exists completely independently of visual stimuli" [9]

In the following sections I describe the technological developments pertaining to both the visual and sonic modalities. First, I look into the history of immersive and augmented visual displays, gestural interaction with 3D environments, and the first experiences in musical interaction design in VR. Second, I propose a reading of the historical developments in digital sound synthesis, spatial audio, virtual agents and AI as constituting a form of 'virtual reality' that pertains only to the auditory modality. I argue that both

[8] Çamcı, Anıl, and Rob Hamilton. "Audio-first VR: New perspectives on musical experiences in virtual environments." *Journal of New Music Research* 49.1 (2020): 1-7.
[9] Turchet, Luca, Rob Hamilton, and Anil Çamci. "Music in extended realities." *IEEE Access* 9 (2021): 15810-15832.

histories are tinkering with similar ideas and concepts, and that this compounded overview can provide insights into the domain of music composition and performance with XR.

1.2 XR technology

1.2.1 Military research

Much of the research and development of XR displays has its origins in military research, with the first HUD (Head-Up Display) implemented in 1940s England during World War II.[10] By the 1950s, military researchers could use HUDs project information into the user's line of sight using a transparent display that would allow pilots to access basic flight information on their windscreens, minimizing their need to glance down and lose sight of the horizon.[11]

In 1961, Philco employees Charles Comeau and James Bryan developed and built the actual first Head-Mounted Display (HMD), *Headsight*. This HMD consisted of a single CRT mounted on a helmet and a magnetic tracking system. The head-tracking was used to remotely control the view of a camera while transmitting the video image to the CRT screen on the helmet. *Headsight* was an early telepresence system used for remote monitoring of dangerous situations and conditions at other locations, with no computer graphics.[12]

In the 1960s, ARPA researcher and Harvard professor, Ivan Sutherland visited the Bell Helicopter Company in Fort Worth, Texas, to see a project involving a remote viewing device on a headset. A servo-controlled camera mapped to the movement of the head, would provide an augmented view of the ground on the pilot's display. The tests he watched at the Bell Helicopter Company, inspired him to think of replacing the camera with a computer-generated world.[13] In 1965 Sutherland published his famous essay 'The Ultimate Display', where he describes the possibilities for computer displays beyond the

[10] Peddie, Jon. "Historical overview: ghosts to real AR to DARPA." *Augmented Reality: Where We Will All Live*. Cham: Springer International Publishing, (2023). 101-133.
[11] Peddie, (2023).
[12] Peddie, (2023).
[13] Peddie, (2023).

visual and auditory modalities. At the end of his essay he describes the 'ultimate display' as something that would almost completely replicate the real world:

> "The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked." [14]

Three years later, in 1968, he demonstrated his prototype for a HMD. Named the *Sword of Damocles*, the prototype hangs from the ceiling, with a mechanical linkage system used to track the movement of the head of the user. It could display transparent wireframe images of cubes, hexagonal molecular shapes, and a 'room' that surrounds the user, all of which could be looked at from different perspectives by means of stereoscopic vision and by calculating six degrees of freedom movement (6DoF) of the user's head (translation and rotation).[15] That same year, Sutherland and his colleague Dave Evans, would go on to start the Evans & Sutherland Computer Corporation, a pioneer in the world of computer graphics.[16]

Throughout the 1970s and 1980s, institutions such as the US Air Force's Armstrong Laboratory, the US Navy, the NASA Ames Research Center, the Massachusetts Institute of Technology, and the University of Carolina at Chapel Hill were doing substantial research on VR and AR.[17] Thomas A. Furness III developed some of the first VR-based prototypes of flight simulators for the US Air Force and continued to work and further develop the cockpit technology for pilots throughout his career.[18] In the 1980s, he was part of the 'Super Cockpit' project, which involved the use of a HMD, head tracking, binaural 3D sound, and gestural input for fighter pilots:

> "The Super Cockpit was envisioned to be a generic crew station which would exploit the natural perceptual, cognitive and psychomotor capabilities of the operator. It is to be

[14] Sutherland, Ivan E. "The ultimate display." *Proceedings of the IFIP Congress*. Vol. 2. No. 506-508. 1965.
[15] Sutherland, Ivan E. "A head-mounted three dimensional display." *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 1968.
[16] Peddie, Jon. "Historical overview: ghosts to real AR to DARPA." *Augmented Reality: Where We Will All Live*. Cham: Springer International Publishing, 2023. 101-133.
[17] Peddie, (2023).
[18] Steinicke, Frank. *Being really virtual*. Immersive natives and the future of virtual reality: Springer, 2016.

based upon several technologies which allow virtual visual, auditory, and tactile worlds to be created for the operator along with an interactive control medium which uses eye, head and hand positions and speech as control inputs." [19]

Throughout the 90s and 2000s, AR technology was continuing to evolve in the defense sector, with demonstrations in combining live AR-equipped vehicles and manned simulators, and the implementation of Battle-field Augmented Reality Systems (BARS).[20]

1.2.2 Creative research

Some researchers and entrepreneurs working with VR were more interested in the creative applications rather than the military ones. In 1960, filmmaker Morton Heilig, patented the first head-mounted display. The patent included small screens to project image towards the eyes, earphones for binaural audio and nozzles to simulate the smell of the virtual environment.[21] He imagined a "cinema for the future" in his failed invention *Sensorama*. An arcade-style cabinet which, as marketed, would provide a multisensory film experience combining 3D stereoscopic vision, binaural audio, scent and vibration.[22]

Over the course of his Ph.D. in computer science at the University of Wisconsin-Madison, Myron Krueger explored telematic interactions between users in a shared 2D virtual space by means of composite video and hand gesture detection.[23] The different interactive works he developed over the course of his Ph.D. would culminate in the development of Videoplace Technology in 1975. This was a virtual reality system that responded with lights and sounds to people's movements and actions without using other interfaces like goggles or gloves.[24] The continuation of his work with this system is described in his famous book *Artificial Reality*.[25]

---

[19] Furness III, Thomas A. "The super cockpit and its human factors challenges." *Proceedings of the human factors society annual meeting*. Vol. 30. No. 1. Sage CA: Los Angeles, CA: SAGE Publications, 1986.
[20] Peddie, Jon. "Historical overview: ghosts to real AR to DARPA." *Augmented Reality: Where We Will All Live*. Cham: Springer International Publishing, 2023. 101-133.
[21] Heilig, Morton L. "Stereoscopic-television apparatus for individual use." U.S. Patent No. 2,955,156. 4 Oct. (1960).
[22] Gutierrez, Nicholaus. "The ballad of morton heilig: on VR's mythic past." *JCMS: Journal of Cinema and Media Studies* 62.3 (2023): 86-106.
[23] Myron W Krueger, "Responsive environments," in *Proceedings of the June 13-16, 1977, national computer conference* (1977), 423–433.
[24] Peddie (2023).
[25] Myron W Krueger, "Artificial reality." (1983).

In 1985, Jaron Lanier and Thomas G. Zimmerman at VPL Research developed a HMD, the *Eyephone*, and a controller, the *Data Glove*.[26] These two devices allowed users to look around and use their hand to interact with objects in VR (the term 'virtual reality' was made popular by Lanier in 1987).[27] In the early 90s Lanier performed 'The Sound of One Hand', likely the first musical performance with VR.[28] He performed an assortment of virtual reality musical instruments of his own creation, the *Rhythm Gimbal* and the *Cybersax,* which he played using the Data Glove.[29] The possibilities he saw for the virtual embodiment of musical instruments through VR are summed up in the following quote:

> "The computer that's running the Virtual Reality will use your body's movements to control whatever body you choose to have in Virtual Reality, which might be human or might be something quite different. You might very well be a mountain range or a galaxy or a pebble on the floor. A piano. . . I've considered being a piano. I'm interested in being musical instruments quite a lot. Also, you can have musical instruments that play reality in all kinds of ways aside from making sound in Virtual Reality. That's another way of describing arbitrary physics. With a saxophone you'll be able to play cities and dancing lights, and you'll be able to play the herding of buffalo's plains made of crystal, and you'll be able to play your own body and change yourself as you play the saxophone. You could become a comet in the sky one moment and then gradually unfold into a spider that's bigger than the planet that looks down at all your friends from high above." [30]

The audience that attended the performance by Lanier, were not immersed in VR and could only see the virtual environment through Lanier's view, which was rendered on a 2D screen.

Researchers from the Electronic Visualization Lab at the University of Chicago were exploring other forms of immersive technology. In 1992, researchers Carolina Cruz-Neira et al., developed the first Cave Automatic Virtual Environment (CAVE) system. Multiple people could be immersed in the same virtual space by projecting the display onto the multiple walls of a cube-shaped physical room. However,

[26] Chuck Blanchard et al., "Reality built for two: a virtual reality tool," in *Proceedings of the 1990 symposium on Interactive 3D graphics* (1990), 35–36.
[27] Peddie, Jon. "Historical overview: ghosts to real AR to DARPA." *Augmented Reality: Where We Will All Live*. Cham: Springer International Publishing, 2023. 101-133.
[28] Jaron Lanier, "The sound of one hand," *Whole earth review* 79 (1993): 30–34.
[29] Jaron Lanier, *Virtual Instrumentation*, Accessed on August 28th, 2024, http://www.jaronlanier.com/instruments.html
[30] Jaron Lanier, *A Vintage Virtual Reality Interview*, Accessed on August 28th, 2023, http://www.jaronlanier.com/vrint.html

the system takes the position of a single user via motion tracking to provide a viewer-centered perspective of the virtual environment that is displayed onto the walls. The goal of such a system was to overcome some of the problems they saw in HMD technology regarding field-of-view, visual acuity, and intrusion.[31] The CAVE system made it difficult for multiple users to interact with the virtual environment. Artists like Julie Martin were developing AR applications for her theater production *Dancing in Cyberspace*, where dancers and acrobats interacted with virtual objects that were projected onto the stage in real-time.[32]

1.2.3 Entertainment industry

In the early 1990s VR started to make its way towards the public through the video game industry. British company W industries (later renamed Virtuality Group PLC) launched their *Virtuality* gaming system in 1990. It consisted of a console in which a user would be standing on the system that ran the software while connected to headsets (or *Visette*) equipped with stereoscopic vision and joysticks, through this system players could interact with each other over a networked system.[33] There was also a seated version of the console, which resembles more a bumper car, or small race kart.

Within the same decade other video game companies such as Nintendo and Sega developed their first VR products, but the commercialization of these systems turned out to be an unsuccessful endeavor.[34] In 1988, Nintendo and Mattel bought the technological design license of the *Data Glove* previously developed by VPL, to develop their next generation controller the *Power Glove*. Although not strictly VR, since it was meant to be used with games that were part of the NES (Nintendo Entertainment System) console, the *Power Glove* pushed forward ideas of gestural and multimodal interaction in video games. Using a combination of photoresistors, conductive ink, magnetic sensors, and an ultrasonic system, the *Power Glove* could track the bending of different fingers and the 3D position of the hand, all while paired

[31] Carolina Cruz-Neira et al., “The CAVE: audio visual experience automatic virtual environment,” *Communications of the ACM* 35, no. 6 (1992): 64–73.
[32] Peddie, Jon. “Historical overview: ghosts to real AR to DARPA.” *Augmented Reality: Where We Will All Live*. Cham: Springer International Publishing, 2023. 101-133.
[33] Steinicke, Frank. *Being really virtual*. Immersive natives and the future of virtual reality: Springer, 2016.
[34] Rustin Webster and Alex Clark, “Turn-key solutions: Virtual reality,” in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 57052 (American Society of Mechanical Engineers, 2015), V01BT02A052.

with button input and a D-Pad.[35]

Nintendo released the first portable 3D stereoscopic vision gaming console, the *Virtual Boy,* in 1995. But it did not have a good reception and did not sell as many units as expected, and its discontinuation followed a year later.[36] Some argue that the reasons for its failure were due to the poor display technology, the isolating gaming experience it created, and a lack of ergonomic design.[37]

In the years to come, the companies in the video game industry moved away from VR, as the technology was 'not there yet', and focused more on handheld controllers and PC games which revolutionized the gaming industry.[38] It was not until 2012 that another attempt would be made to commercially release another VR product. After a year-long kickstarter campaign, the company Oculus released their first HMD: the Rift DK1.[39] It proved to be a major advancement compared to all the previous attempts at releasing VR products by the video game industry, and while it still had some issues regarding display and position tracking, subsequent hardware and software releases followed to solve many of these issues.[40]

1.2.4 Current technologies

Alongside the development of VR and CAVE-like technology, different tools for full body tracking, and hand gesture tracking are used to create virtual representations of our bodies and can be applied for motion, and gestural interaction with virtual environments. Infrared sensor motion capture technology has been widely used in film and video game animation,[41] sports performance,[42] and other applications. More

[35] Entertainment, Abraham Gentile. "Power Glove." (1989).
[36] Steven Boyer, "A virtual failure: Evaluating the success of Nintendo's Virtual Boy," *The Velvet Light Trap*, no. 64 (2009): 23–33.
[37] Zachara, Matt, and José P. Zagal. "Challenges for success in stereo gaming: a Virtual Boy case study." *Proceedings of the International Conference on advances in computer entertainment technology*. 2009.
[38] Williams, Dmitri. "Structure and competition in the US home video game industry." *International Journal on Media Management* 4.1 (2002): 41-54.
[39] Gleasure, Rob, and Joseph Feller. "A rift in the ground: Theorizing the evolution of anchor values in crowdfunding communities through the oculus rift case study." *Journal of the Association for Information Systems* 17.10 (2016): 1.
[40] Parth Rajesh Desai et al., "A review paper on oculus rift-a virtual reality headset," *arXiv preprint arXiv:1408.1173*, (2014).
[41] Bregler, Chris. "Motion capture technology for entertainment [in the spotlight]." *IEEE Signal Processing Magazine* 24.6 (2007): 160-158.
[42] Ortega, Basilio Pueo, and José Manuel Jiménez Olmedo. "Application of motion capture technology for sport performance analysis." *Retos: nuevas tendencias en educación física, deporte y recreación* 32 (2017): 241-247.

portable systems have also been developed in the last decade. The Microsoft Kinect was originally created for the Xbox One console but Microsoft made a developer version available in 2012, which allowed users to adapt the Kinect's body tracking capabilities for all sorts of applications, from scientific to art.[43] [44] [45] The LEAP motion controller developed in the 2010s is another portable computer vision system dedicated to hand motion tracking and gesture recognition. It has been used in multiple VR and AR applications.[46] [47]

Through cameras on mobile computing devices, Augmented Reality (AR) technology can render virtual elements and insert them in real world environments through markers.[48] There is also an increasing interest to interface VR and AR with web development, enabling users to browse and interact with the web through such technologies.[49] Since then, we have seen a wide development of commercially available headsets from different companies, with applications of VR and AR systems for professional training,[50] entertainment, military training, therapy,[51] and scientific research.[52]

---

[43] Zerpa, Carlos, et al. "The use of microsoft Kinect for human movement analysis." *International journal of sports science* 5.4 (2015): 120-127.
[44] Andersen, Michael Riis, et al. "Kinect depth sensor evaluation for computer vision applications." *Aarhus University* (2012): 1-37.
[45] Graham-Knight, Kimberlee, and George Tzanetakis. "Adaptive music technology using the Kinect." *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 2015.
[46] Wozniak, Peter, et al. "Possible applications of the LEAP motion controller for more interactive simulated experiments in augmented or virtual reality." *Optics Education and Outreach IV*. Vol. 9946. SPIE, 2016.
[47] Păvăloiu, Ionel-Bujorel. "Leap motion technology in learning." *Edu world 7th international conference*. 2017.
[48] Young-geun Kim and Won-jung Kim, "Implementation of augmented reality system for smart- phone advertisements," *international journal of multimedia and ubiquitous engineering* 9, no. 2 (2014): 385–392.
[49] Xiuquan Qiao et al., "Web AR: A promising future for mobile augmented reality—State of the art, challenges, and insights," *Proceedings of the IEEE* 107, no. 4 (2019): 651–666.
[50] Renganayagalu, Sathiya Kumar, Steven C. Mallam, and Salman Nazir. "Effectiveness of VR head mounted displays in professional training: A systematic review." *Technology, Knowledge and Learning* (2021): 1-43.
[51] Bowman, Doug A., and Ryan P. McMahan. "Virtual reality: how much immersion is enough?." *Computer* 40.7 (2007): 36-43.
[52] Diederick C Niehorster, Li Li, and Markus Lappe, "The accuracy and precision of position and orientation tracking in the HTC vive virtual reality system for scientific research," *i-Perception* 8, no. 3 (2017): 2041669517708205.

1.3 Sonic Simulations: Disembodied Sound

> *"Much of the work that many of us have done in electronic music invests heavily in the idea of hearing recording as a primary experience, and in a world in which we peer at a kind of virtual reality through the windows, or lenses, of loudspeakers"*
>
> - *Paul Lansky (The Importance of Being Digital*, *TART Foundation and Gaudeamus Foundation*)

In this section I describe landmark developments in computer music that could be considered to some extent as part of the history of XR (at least for composers and creatives working with this medium). I argue that music researchers, composers and creatives have been developing ideas of 'illusion', 'simulation' and 'virtual worlds' through the shaping and sculpting of 'disembodied' sound (recorded and digital sound), curating the sonic reality of audiences for decades, in a version of 'virtual reality' that concerns only the auditory modality.

It is not my intention here to review the complete history of digital sound synthesis, but to understand it as a tool for simulating reality in a purely sonic dimension – in contrast to the development of immersive technologies with a heavy inclination towards the visual modality. I have selected examples and ideas from digital sound synthesis that resonate with the development of XR throughout the 1960s, 70s and 80s – creating distinct 'sonic worlds', blending the real and the 'virtual' in music composition, and digital sound synthesis as allowing us to sculpt a sonic reality to our own desires – aesthetically, these developments offered uncharted territory for composers in computer music, with many exploring this new space where real world sounds and synthetic sounds could interact with each other.

1.3.1 Musique Concrète and Elektronische Musik

Before the advent of computers and digital sound synthesis, composers working at different radio laboratories across Europe were pushing the boundaries of electroacoustic music composition using analog hardware. Two of the prominent schools of thought that emerged from the technological developments of the post-WWII era were those of *musique concrète*, in Paris, and *elektronische Musik*, in Cologne. At RTF (Radiodiffusion Télévision Française), Pierre Schaffer was experimenting with tape recordings of everyday sounds, while developing his ideas on the *objet sonore*, reduced listening and the syntax of *musique concrète*.[53] At the NWDR (Norwestdeutscher Rundfunk) in Cologne, composers such as Karlheinz Stockhausen, Herbert Eimert and Robert Beyer, were experimenting and applying the musical concepts from music serialism to the different parameters in analog synthesis in what came to be known as *elektronische Musik*, or simply ‘electronic music’.[54]

Meanwhile in the US, in the early 1950s, composers such as Louis and Bebe Barron together with John Cage, Earle Brown, David Tudor, Morton Feldman and Christian Wolff, began to experiment with making music directly on to tape. Halfway through the decade composers Otto Luening, Vladimir Ussachevsky and Milton Babbitt were starting to petition and assemble electronic music laboratories at Columbia and Princeton in the US.[55] These composers soon learned that there was also important research regarding the analysis and synthesis of sound happening at Bell Telephone Laboratories under the direction of Max Matthews, who would pioneer digital sound synthesis.

1.3.2 Digital Sound Synthesis

In 1957, Matthews published the first version of MUSIC, a programming language with which he could use computers to synthesize the sound of different waveforms. The program went through many iterations in the following decades to eventually become CSOUND in 1986.[56] The first version of this program,

[53] Manning, Peter. *Electronic and computer music*. Oxford University Press, USA (2013), page 20.
[54] Manning (2013), page 40.
[55] Manning (2013), page 74.
[56] Manning (2013), page 190.

MUSIC I, was installed in an IBM 704, a first generation computer, and it could only synthesize triangle waveforms. The arrival of transistor-based circuits introduced a new family of machines in the late 1950s. With more efficient computer architecture that replaced the unreliable vacuum-tube logic circuitry valve, second generation computers such as the IBM 7094 were used to perform complex tasks such as synthesizing a singing voice for the first time.[57] In 1961, John L. Kelly Jr and Carol Lockbaum programmed the computer to sing the song *Daisy Bell* with the accompaniment written by Matthews.[58] Matthews digitally synthesized the sounds using an early version of his MUSIC software.[59] The song is also known as *Bicycle built for two*. A historical moment that would later be mediated and introduced to popular media through Stanley Kubrick's science fiction movie *2001: A Space Odyssey* in 1968, with the iconic scene where the AI computer HAL9000 sings *Daisy Bell* as it is being deactivated.[60]

The first digital synthesis techniques programmed by Matthews using his MUSIC-N series of languages consisted of multiple 'instrument units', with which he could address oscillators with wavetables (i.e. sine, triangle, square), and could arrange in a process of additive synthesis to generate new sounds, using the computer as a musical instrument.[61]

Jean-Claude Risset joined the research team at Bell Labs in 1964, where he investigated the timbrical characteristics of trumpet by performing spectral and amplitude analysis on recordings of a professional trumpet player. He would use the data from the analysis to generate synthetic brass sounds.[62]

Influenced by J.J. Gibson's theory of perception, the researchers at Bell Labs understood the need to apply this 'ecological'[63] view and study the psychoacoustic processes involved in the perception of sounds by listeners.[64] The knowledge of the acoustic properties of sounds in conjunction with studies in psychoacoustics would bring forward the creation of realistic digital representations of sound. Although

---

[57] Manning (2013), page 188.
[58] The IBM 7094 is The First Computer to Sing, https://www.historyofinformation.com/detail.php?entryid=4445 accessed November 20, 2025
[59] Manning (2013), page 193.
[60] https://www.youtube.com/watch?v=c8N72t7aScY accessed November 13, 2025
[61] Mathews, Max V. "The Digital Computer as a Musical Instrument: A computer can be programmed to play" instrumental" music, to aid the composer, or to compose unaided." *Science* 142.3592 (1963): 553-557
[62] Risset, Jean-Claude. "Computer study of trumpet tones." *The Journal of the Acoustical Society of America* 38.5_Supplement (1965): 912-912.
[63] Gibson, James J. "Theories of Perception." (1951).
[64] Mathews (1963).

the results of that time were far from achieving the complexity of contemporary sound synthesis tools, sometimes perceived as dull and 'electronic',[65] the possibility itself of digitally synthesizing sounds and shaping their timbre, would spark provocative thoughts in some composers.

> "Synthesis creates a 'virtual' world of sound - a purely sonic world, without a visible physical counterpart. (...) A virtual world of sound is one which has the virtue to suggest a different reality, an immaterial, illusory world, often invisible, anchored in our perception rather than in our environment. Moreover synthetic sounds can evoke elements of our familiar physical world: our perception has a strong tendency to assimilate the unknown to the familiar." [66]

In this opening paragraph of Risset's article, what makes synthesized sound have the character of 'virtual' for him, derives from the 'disembodiment' of sound from its mechanical production in the physical world. Nonetheless, synthesized sounds are capable of evoking a sense of familiarity to us – but this is grounded in our perception, and in our own knowledge of sounds in the physical world. This definition of 'virtual' is useful to also understand previous developments in *musique concrète* and *elektronische musik*, where sound is also severed from physically actuated mechanisms of production.

In another excerpt Risset uses the word *simulacra* to describe the synthesis of sounds that he used in his works to imitate musical instruments or the human voice.[67] The use of the word 'simulacra' is interesting as it denotes the quality of early sound synthesis: a sound we can potentially ascribe an identifiable source, a familiar sound that approximates reality but not a perfect replica.

Together with the developments from Matthews and Risset, other synthesis techniques capable of simulating a wider variety of familiar or 'natural'[68] sounds would also be developed in the 1970s and 1980s. John Chowning, in 1973, published his work on a new powerful tool for sound synthesis using frequency modulation (FM). FM was originally applied for radio transmission of signals, but in FM sound synthesis, carrier and modulating frequencies are within the hearing range, using the side bands to

[65] Risset, Jean-Claude. "Real-world sounds and simulacra in my computer music." *Contemporary music review* 15.1-2 (1996): 29-47.
[66] Risset (1996).
[67] Risset (1996).
[68] Sounds that are physically actuated in the real world.

generate complex spectra.[69] One of the advantages of FM is its capacity to create both harmonic and inharmonic spectra which lends itself to simulate a wide range of natural sounds, most notably bells and drums.

These developments were pushing the boundaries of what could be perceived – from a sonic perspective – as synthetic and what could be perceived as real, and how these opposite ends could merge or diverge in music composition, as Risset mentions in the following excerpt:

> "I have tried to synthesize a lively world of sounds, distinct from the acoustic world – the world of real objects – and to control synthesis so that these two worlds can occasionally merge as well as diverge and contrast." [70]

In this context, what constitutes 'the world of real objects' for Risset is the perception of the causality of sound. For him, if someone is able to associate a form of mechanical production with the synthesized sound, that gives it a stronger identity making it be perceived as real.

> "Acoustic sounds are constrained by the way they are generated. We attribute a clear identity to sounds when we can guess which mechanical process produced them - hitting, scraping, bowing. In contradistinction, the generation of digital sounds is devoid of material constraints: one can thus shape or sculpt sound in arbitrary ways." [71]

This back and forth between what is perceived as an illusion and perceived as real, is what constitutes for Risset the creative space for composers working with digital sound synthesis at the time. He emphasizes this aspect in Chowning's composition, *Turenas:*[72]

> "*Turenas* carves its own space through the trajectories of illusory sound sources – but the timbres are also given trajectories that take them from harmonic to inharmonic and back

[69] Chowning, John M. "The synthesis of complex audio spectra by means of frequency modulation." *Journal of the audio engineering society* 21.7 (1973): 526-534.
[70] Risset, Jean-Claude. "Real-world sounds and simulacra in my computer music." *Contemporary music review* 15.1-2 (1996): 29-47.
[71] Risset (1996).
[72] Turenas is an anagram for 'natures', in Manning, Peter. *Electronic and computer music*. Oxford University Press, USA, 2013, page 195.

to harmonic spectra, occasionally evoking either closely or remotely familiar sounds – birds, bells, drums – but with a ductility which is the mark of synthesis."[73]

In other examples, we find different relationships between the synthetic and the real, or illusion and real. For example, Charles Dodge and Paul Lansky were both working with voice and speech material in their compositions. Dodge was using Linear Predictive Coding (LPC), an analysis-synthesis technique, to generate artificial voices.He was applying methods from musical serialism to the manipulation of synthesized speech.[74] In his 'Speech Songs', Dodge's work does not try to convince us of how realistic synthetic the artificial voice can be, but rather it emphasizes how unreal it is through the manipulation and modulations in speed and pitch. It is in these unrealistic manipulations that the 'humanness' of the voice appears to us, where we can ascribe an emotion or tone in our attempt to understand the ramblings of the synthetic voice – sometimes it sounds like it is asking a question, or it sounds drunk, or angry).

In contrast, in the work of Lansky, we find a desire to transform the 'unmusical' sounds of the real world into 'musical' sounds. A pursuit similar to that of *musique concrète*, with the caveat that this would be achieved by means of reduced listening. Lansky's approach is that of hybridization, to enhance and extend the musical potential of everyday sound. For example, in *Six Fantasies on a Poem by Thomas Campion*, Lansky musicalizes the consonants and vowels of speech using a combination of LPC and filtering, in what is known as a Vocoder. Similarly, in *Night Traffic*, the traffic sounds are augmented by the chords they very produce in a Vocoder. The piece transits from the very clear traffic sounds to a sound world of chords while never losing its relationship to the traffic sounds, in a beautiful piece that creates illusions and augmented versions of real world sounds.

1.3.3 Physical Modeling

Following the idea of a sound being perceived as 'realistic' stemming from the association of the sound to its mechanical production in the physical world, researchers explored how to simulate the physical

[73] Risset, Jean-Claude. "Real-world sounds and simulacra in my computer music." *Contemporary music review* 15.1-2 (1996): 29-47.
[74] Manning, Peter. *Electronic and computer music*. Oxford University Press, USA, (2013).

mechanisms that produce sound to implement it in digital sound synthesis.[75] In 1971, Hiller and Ruiz published their research on the solutions to differential equations to describe the oscillations of vibrating objects, which they used to generate sounds on a computer.[76] At the time, this was a computationally expensive process, but with the expansion of the microprocessor market towards the 1980s, computers were better equipped to better perform demanding tasks such as physical modeling.[77]

Other researchers found simpler and more efficient solutions for physical modeling, as is the case with the karplus-strong algorithm, developed by Kevin Karplus and Alex Strong.[78] It uses a simple delay line, noise and feedback to synthesize the sounds of strings and drums.

Claude Cadoz and his group ACROE, analyzed musical instruments in terms of excitation, and vibration structures. They looked into the different forms of excitation on musical instruments and their relation to vibrating strings, membranes and air columns.[79]

1.3.4 Granular synthesis: simulating environmental textures

In the 1980s, other forms of analyzing sound would bring along new forms of synthesis with composers thinking about 'virtual' sound in much more malleable terms, applying stochastic and environmental models to generate sounds and compositions.

In his theory of communication, Dennis Gabor refers to the smallest components of sounds as *acoustic quanta* or 'grains', and how these can be used to represent any sound.[80] Xenakis first theorized about the compositional possibilities of such a system if it were to be programmed in a computer capable of managing large quantities of grains.[81]

---

[75] Risset, Jean-Claude. "Real-world sounds and simulacra in my computer music." *Contemporary music review* 15.1-2 (1996): 29-47.
[76] Hiller, Lejaren, and Pierre Ruiz. "Synthesizing musical sounds by solving the wave equation for vibrating objects: Part 1." *Journal of the Audio Engineering Society* 19.6 (1971): 462-470.
[77] Manning, Peter. *Electronic and computer music*. Oxford University Press, USA, (2013).
[78] Karplus, Kevin, and Alex Strong. "Digital synthesis of plucked-string and drum timbres." *Computer Music Journal* 7.2 (1983): 43-55.
[79] Cadoz, Claude, et al. "Responsive input devices and sound synthesis by stimulation of instrumental mechanisms: The cordis system." *Computer music journal* 8.3 (1984): 60-73.
[80] Gabor, Dennis. "Theory of communication. Part 1: The analysis of information." *Journal of the Institution of Electrical Engineers-part III: radio and communication engineering* 93.26 (1946): 429-441.
[81] Xenakis, Iannis. *Formalized music: thought and mathematics in composition*. No. 6. Pendragon Press, (1992).

Curtis Roads made the first implementations of granular synthesis using digital computers in 1975 and 1981. He would generate grains from sampled percussion and saxophone sounds using an elaborate process of punched paper cards that were fed to a computer to calculate the different values for each 'grain', the resulting waveforms were later written onto magnetic tape for playback. Roads would organize the grains into *events*, which are characterized by containing a set of parameters that indicate the beginning time, duration, pitch, and amplitude, among other parameters. These *events* could be plotted in different two-dimensional shapes on a frequency-versus-time graph to control large quantities of grains and generate uniform textures and clouds of evolving sound spectra.[82] Granular synthesis allowed for novel forms of manipulation of recorded sound that contrasted with approaches and ideas developed in *musique concrète*. One of the advantages of this technique is the possibility to alter the duration of a signal while preserving the frequency/pitch content, and vice versa, change the frequency/pitch content of a signal while retaining its original duration.

The advancements made by Roads were echoed by his contemporary Barry Truax in the development of real-time granular synthesis using a DMX-1000 Digital Signal Processor.[83] He explored the use of additive synthesis, FM synthesis, and recorded sound in the form of grains or sound fragments for real-time performance and composition. He combined different control strategies to compose and perform pieces through the use of *presets*, *ramps* and *tendency masks*.[84] These allowed him to determine the large-scale form and macro-level texture of his compositions. Truax had been interested in stochastic procedures to generate and control audio spectra since the 1970s, and had been working on using probabilistic models for his POD (POisson Distribution) system for music composition.[85] Based on a perceptual study by John McKay, Truax understood that high-density sonic events or granular sonic textures can create the impression of 'flows', 'swarms', 'textural bands' and 'massed sonorities' in high-density sonic events.[86] The acoustic results from granular synthesis could simulate the inner

[82] Roads, Curtis. "Introduction to granular synthesis." *Computer Music Journal* 12.2 (1988): 11-13.
[83] Wallraff, Dean. "The DMX-1000 signal processing computer." *Computer Music Journal* (1979): 44-49.
[84] Truax, Barry. "Real-time granular synthesis with a digital signal processor." *Computer Music Journal* 12.2 (1988): 14-26.
[85] Manning, Peter. *Electronic and computer music*. Oxford University Press, USA, (2013). page 205
[86] MacKay, John. "On the perception of density and stratification in granular sonic textures: An exploratory study." *Journal of New Music Research* 13.4 (1984): 171-186.

complexity and the statistical nature (randomness) of environmental sound. Truax's piece *Riverrun* (1986) was created using real-time granular synthesis and later mixed for an octophonic sound system. The piece opens up with the individual accumulation of 'droplets' (in this case, grains of sound) which gradually transform into "a sound environment in which stasis and flux, solidity and movement coexist in a dynamic balance similar to a river, which is always moving yet seemingly permanent".[87]

1.4. Sonic Simulations: Spatial Audio, simulating space and movement

Our perception of sound does not only comprehend its spectral and morphological qualities, but also its location and movement, revealing the qualities of the space through which it travels. Spatial audio is one of the fundamental ways in which XR experiences are reinforced through sound.

1.4.1 Early experiments in recording and playback

Advancements in tape recording (as used in *musique concrète*) and sound synthesis (as used in *elektronische musik*) technology led to the severance of sound its mechanical mode of production, the speaker became the new embodiment for the playback of tape and sound synthesis. Composers started to think of the speaker as another material with which to compose music, questioning its spatial arrangement to convey 'sound images' and thinking about it in instrumental ways (feedback). At the RTF in Paris in the 1950s, engineer Jacques Poullin developed his *potentiomètre d'espace* for Pierre Schaeffer. A hand-held device that used gestural control to distribute a tape playback signal between multiple loudspeakers, composing the localization of tape sounds in real-time.[88] Composers in Cologne were exploring different microphone and speaker configurations for recording quadraphonic sound in the late 1950s.[89] Stockhausen premiered *Gesang Der Jünglinge* with a five-speaker configuration for the playback in 1956.[90] Edgard Varése, in collaboration with Le Corbusier, composed *Poème électronique* for the Philips Pavilion in the 1958 World Fair in Brussels. The piece was played back in a multi-channel system

---

[87] Truax, Barry. "Real-time granular synthesis with a digital signal processor." *Computer Music Journal* 12.2 (1988): 14-26.
[88] Manning, Peter. *Electronic and computer music*. Oxford University Press, USA (2013), page 27.
[89] Manning (2013), page 62.
[90] Lyon, Eric. "Spatial orchestration." *Proceedings of the 5th Sound and Music Computing Conference*. 2008.

of several hundred loudspeakers throughout the architectural space, creating an immersive audiovisual experience for the attendees.[91]

1.4.2 Multi-channel spaces and spatial audio formats

Multiple institutions began to implement multi-channel speaker arrays in their concert halls and electronic music laboratories. The Groupe de Recherches Musicales (GRM) in France completed their Acousmonium in 1974, a concert hall with eighty speakers distributed across the room. Up to 48 channels of sound could be diffused with their system.[92] The Birmingham Electro-Acoustic Sound Theatre (BEAST) was established in 1982 by Jonty Harrison, and was modeled after the Acousmonium in terms of spatial arrangement of speakers for diffusion, with the aim to liberate the listening experience from the 'sweet spot'.[93] Nowadays it is fairly common for performance spaces in academic institutions to have multi-speaker audio configurations for music performance.

Alongside the creation of spaces with multi-channel performances, researchers had been developing different techniques for spatial audio. The Ambisonics format was first developed throughout the 1970s, with research involving the capture,[94] playback[95] and the psychoacoustics[96] involved in 3D sound.

Developments in binaural audio technology helped determine the filtering and delay produced by the head and ear shape of a person, and its relation to the perception of localized sound in space, in what is called Head Related Transfer Function (HRTF).[97] Headphone playback systems are most commonly used with HRTFs to digitally render sound images that simulate the position of a sound source in 3D space.[98]

Towards the end of the 1990s, researchers at IRCAM developed the *Spatialisateur* or SPAT a

---

[91] Mondloch, Katie. "A Symphony of Sensations in the Spectator: Le Corbusier's Poeme electronique and the Historicization of New Media Arts." *Leonardo* 37.1 (2004): 57-62.
[92] Manning, Peter. *Electronic and computer music*. Oxford University Press, USA (2013), page 445.
[93] Knight-Hill, Andrew. "Theatres of sounds: The role of context in the presentation of electroacoustic music." *Scene* 6.2 (2018): 165-175.
[94] Smith, J. HOWARD, and K. JH. "The sound field microphone." *db Magazine* 12.7 (1978).
[95] Gerzon, Michael A. "Periphony: With-height sound reproduction." *Journal of the audio engineering society* 21.1 (1973): 2-10.
[96] Gerzon, Michael A. "Surround-sound psychoacoustics." *Wireless World* 80.1468 (1974): 483-486.
[97] Møller, Henrik. "Fundamentals of binaural technology." *Applied acoustics* 36.3-4 (1992): 171-218.
[98] Li, Song, and Jürgen Peissig. "Measurement of head-related transfer functions: A review." *Applied Sciences* 10.14 (2020): 5014.

software for spatialization of sound, a tool meant to be integrated directly in the process of composing with digital sound synthesis and signal processing.[99]

Other methods such as Vector Based Amplitude Panning (VBAP), developed by Ville Pulkki at Aalto University in Finland, had the benefit of being adaptable to a variable number of speakers, making it applicable to a variety of performance spaces.[100]

Different approaches involved the development of specific hardware arrangements, such is the case with Wave Field Synthesis which uses loudspeaker linear arrays that are equally spaced, to simulate wave fronts properties of individual sound sources.[101]

In the XR literature, immersion is a concept that is used to describe the technology involved in displaying computer generated graphics.[102] For example, a HMD is more immersive than a 2D screen; a CAVE system can immerse multiple people at the same time, etc. In the case of multi-channel speaker systems and binaural audio, these are both immersive technologies.

The developments in spatial audio recording technology, digital signal processing and multichannel speaker systems present not only new possibilities for music composition, but have implications on how we make meaning of recorded and synthesized music. In the words of composer Paul Lansky:

> "In essence then, a recording can create what could reasonably (although unfortunately) be called a virtual world and we as listeners have become acculturated to peering into that world, accepting and disregarding its limitations and its contradictions of reality. (...) the wonders of stereo reproduction and now the marvel of multi-channel digital sound poke at the potential for recorded sound to ultimately be indistinguishable from the real thing. (Whether or not this is ever possible is beside the point. We certainly are approaching that goal). In each case we formerly peered into a world that had some sort of curtain around it, and was only a weak approximation of reality as we know it. As the technology

[99] Manning (2013), page 448.

[100] Pulkki, Ville. "Virtual sound source positioning using vector base amplitude panning." *Journal of the audio engineering society* 45.6 (1997): 456-466.

[101] Berkhout, Augustinus J., Diemer de Vries, and Peter Vogel. "Acoustic control by wave field synthesis." *The Journal of the Acoustical Society of America* 93.5 (1993): 2764-2778.

[102] Mel Slater and Sylvia Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," *Presence: Teleoperators & Virtual Environments* 6, no. 6 (1997): 603–616.

improves the curtain becomes more transparent. But rather than try to cast recording as an incomplete representation of reality it is more useful today to imagine that there are two realities, the experience of recorded sound and the experience of live sound." [103]

The technologies involving spatial audio continue to improve on every front with multiple techniques using arrays of microphones,[104] dummy heads,[105] and ambisonic microphones for recording.[106] With the technological improvements in virtual room acoustics, binaural audio, personalized HRTFs and 6-DoF immersive sound, will composers in the 21st century try to make recorded sound indistinguishable from the experience of live sound? or should they, as Lansky imagines, intend to keep both as different realities for audiences to experience?

1.5 Sonic Simulations: Musical Agents

Digital sound synthesis gives composers the capacity to create virtual sonic worlds of their own, whether if it is in real-time or through an iterative process of generation, manipulation and editing. Humans are the ones carrying out these processes, making the decisions through notation, programming and performing with various systems. If a computer is given the task to perform these decisions, or to perform live music together with other agents (human and non-human), could it constitute a form of 'virtual reality'?

1.5.1 Networks and computer players

With the arrival of the microprocessor in the 1970s, some composers were interested in treating the computer as another musician they could perform with or as a musical actor on its own.[107] The League of Automatic Composers for example, were invested in creating networks of computers that would interact with each other to generate sounds, with an approach to the computer network as one large, interactive musical instrument.[108]

[103] Lansky, Paul. *The Importance of Being Digital*. TART Foundation and Gaudeamus Foundation, 2004, page 8.
[104] Politis, Archontis. "Microphone array processing for parametric spatial audio techniques." (2016).
[105] Vorländer, Michael. "Past, present and future of dummy heads." *Proc. Acústica, Guimarães, Portugal* (2004): 13-17.
[106] Zotter, Franz, and Matthias Frank. *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature (2019).
[107] Manning, Peter. *Electronic and computer music*. Oxford University Press, USA, (2013), page 219.
[108] Perkis, Tim, et al. *The League of Automatic Music Composers, 1978-1983*. New World Records, 2007.

In the 1980s, Barry Vercoe and Larry Beauregard were developing a *synthetic performer* to accompany a live musician.[109] The computer program would follow the score that was being played by a performer.[110] In 1989 Jean-Claude Risset composed *duet for one pianist*, in this piece a pianist and a computer play together on a single Yamaha Disklavier. The key presses by the pianist are sent via MIDI to a Macintosh computer where a program determines in what way the computer will respond, sending back MIDI signals to trigger the keys on the Disklavier.[111]

1.5.2 George E. Lewis and Voyager

What is embedded in the previous examples is the idea to 'synthesize' or 'virtualize' a musical agent capable of performing alongside human performers by making its own decisions and/or responding to what other performers are playing. One of the pioneers in this matter is composer, improviser and trombonist George E. Lewis. He gathered an interest in interactive systems and computer music after meeting David Behrman on a visit to Mills College in California in 1977.[112] While there, he also had the chance to see a performance by the League of Automatic Music Composers. Their use of interconnected KIM microprocessors to generate sounds in automatic (yet random) ways made an impression on Lewis, to him "it sounded a lot like a band of improvising musicians".[113] After this experience he acquired a KIM-1 for himself and started to learn how to program assembly language and how to make cheap digital-to-analog converters to get his own sounds into the computer.[114] He then premiered his first piece of interactive computer music, *The KIM and I*, at the Kitchen performance space in downtown New York in 1979.[115] This was an interactive system in which he would control a Moog synthesizer by playing his trombone, using his custom-built computer with the KIM-1 microprocessor.

In 1984, while at IRCAM he continued to develop his interactive computer system and premiered

[109] Vercoe, Barry. "The synthetic performer in the context of live performance." *Proceedings of International Computer Music Conference*. 1984.
[110] Risset, Jean-Claude. "Computer music: why." *Internet Proceedings of Composers' Forum Austin*. 2003.
[111] Risset (2003).
[112] Steinbeck, Paul. "George Lewis's voyager." *The Routledge companion to Jazz studies*. Routledge, 2018. 261-270.
[113] Roads, Curtis. "Improvisation with George Lewis." *Composers and the Computer* (1985): 75-88.
[114] Roads (1985).
[115] Lewis, George E. "Living with Creative Machines: An Improvisor Reflects." *AfroGEEKS: Beyond the Digital Divide* (2007): 83-99.

Rainbow Family. In this piece, three DX-7 synthesizers were controlled by three Apple II computers – each running the custom software programmed by Lewis. The performances were between the three computers and four human improvisors, including contrabassist Joelle Leandre, saxophonist Steve Lacy, bass clarinetist Douglas Ewart, and guitarist Derek Bailey. Lewis took inspiration in AI, cybernetics, and free improvisation practices to position the 'creative machines' as central actors in *Rainbow Family*.[116] Each of the computers in this network is considered by Lewis as a separate improviser, each making its own performance decisions. Lewis strived for a 'sonic individuation' to be perceived for each computer. An individuation that would derive from the decision making processes of each computer to develop a unique 'sound'.[117] He used separate speakers and limited the sets of sounds available for the output of each machine in an attempt to enhance the individuation through spatial and sonic means, but in the end the system was still perceived by performers and audience as a whole 'unitary' machine.[118] *Rainbow Family* would form the foundations for what would later become *Voyager*.[119]

The first iterations of *Voyager* were programmed between 1986 and 1988, while Lewis was at the Studio for Elektro-Instrumentale Muziek (STEIM) in Amsterdam. He used the Forth programming language and a similar setup to that of *Rainbow Family*, with a Macintosh connected to Yamaha synthesizers.[120] During the 1990s it was updated to incorporate MIDI, and in the 2000s Lewis decided to recreate it entirely in Max/MSP,[121] which allowed *Voyager* to interface with more instruments, such as MIDI-capable acoustic pianos.[122] Even though it was ported to different programming environments and connected to different instruments, *Voyager* retained the functionality and aesthetics that characterized it throughout all of its versions. It is conceived not as a single improviser, but rather as a 'virtual improvising orchestra' modeled after a Javanese gamelan ensemble, where the control of the musical

[116] Lewis, George. "Co-creation: early steps and future prospects." *In Bernard Lubat, Gérard Assayag, Marc Chemillier. Artisticiel/Cyber-Improvisations. Phonofaune, 2021, Dialogiques d'Uzeste.* (2021).
[117] Lewis, George E. "Improvised music after 1950: Afrological and Eurological perspectives." *Black music research journal* (1996): 91-122.
[118] Lewis (2021).
[119] Lewis, George E. "Too many notes: Computers, complexity and culture in voyager." *Leonardo music journal* 10 (2000): 33-39.
[120] Lewis (2000).
[121] https://cycling74.com/articles/an-interview-with-george-lewis-and-damon-holzborn-part-1 accessed December 3rd, 2025
[122] Steinbeck, Paul. "George Lewis's voyager." *The Routledge companion to Jazz studies*. Routledge, 2018. 261-270.

process is shared among players without the need of a central authority. This model consists of 64 MIDI-controlled players or voices that run asynchronously, capable of generating music in real time.[123]

For Lewis it is important that the machine is capable of performing its own ‘voice’, for it to pursue expression without it being tied to direct real-time responses to a human performer. *Voyager* is described by Lewis as a ‘player’ program, “where the computer system does not function as an instrument to be controlled by a performer.”[124] He borrows this definition from Robert Rowe’s taxonomy of ‘player’ and ‘instrument’. For Lewis, in a performance of *Voyager*, the computer program and the human performer play together on equal terms, involving “parallel streams of music generation, emanating from both the computers and the humans – a nonhierarchical, improvisational, subject-subject model of discourse, rather than a stimulus/response setup”.[125] According to Lewis, *Voyager* is not simply a ‘virtual improvising orchestra’, but rather the virtual embodiment of African-American cultural practice through a virtual agent.[126] For him, the creation of musical computer software reflect the ideas and world view of its creators:

> “Musical computer programs, like any texts, are not “objective” or “universal,” but instead represent the particular ideas of their creators. As notions about the nature and function of music become embedded into the structure of software-based musical systems and compositions, interactions with these systems tend to reveal characteristics of the community of thought and culture that produced them” [127]

And so for him the creation of musical virtual agents does not involve only the musical and technical challenges involved in giving the system a musical ‘personality’ and the capacity to be present in the interact with human performers, but also requires us to examine the sociological and cultural assumptions that go into the development of music technology on a broader scale.

In the XR literature, it is important for immersive experiences to create a sense of ‘presence’ in

---

[123] Lewis, George E. “Too many notes: Computers, complexity and culture in voyager.” *Leonardo music journal* 10 (2000): 33-39.
[124] Lewis (2000).
[125] Lewis (2000).
[126] Lewis, George E. “Interacting with latter-day musical automata.” *Contemporary Music Review* 18.3 (1999): 99-112.
[127] Lewis (2000).

the users.[128] This is achieved by creating the illusion of 'being there' by using visual and auditory stimuli in ways that do not break the illusion of presence. One could argue that *Voyager* constitutes a form of 'virtual reality' by creating a sense of 'presence' for other performers interacting with it, and for the audience watching a performance. In its capacity to play by itself, or to decide to interact with the sonic material that it is 'hearing', its responses are all plausible, and create the illusion that a musical agent is 'present' performing with other humans.

1.6 Sonic Simulations: Artificial Intelligence

In *Rainbow Family* and *Voyager*, Lewis was programming complex tasks that involved listening, interpreting, and deciding, based on what was currently happening (or had happened). He programmed the way the computer listens based on his own intuition, with a strong basis and guidance from his own ear as an experienced and seasoned improvisor. Many of the problems that he was trying to solve in his interactive systems were being researched by people in the field of Artificial Intelligence (AI).

1.6.1 Generative models and the problem of digital representation in music

Curtis Roads, in his paper on *Artificial Intelligence and Music,* published in 1980, describes multiple attempts from researchers since the late 1960s to systematize, encode, parse and formalize musical patterns through language recognition systems in programming languages such as Lisp.[129] Most of them were based on principles that are most relevant to western classical music such as tonality, rhythm, melody, and harmony.[130] [131] He also describes multiple examples of early generative music models that focused on the simulation and modeling of traditional music, folk tunes, gregorian chant, medieval polyphony, Bach counterpoint, sonata-form structures, jazz improvisation, figured bass and melody

[128] Mel Slater and Sylvia Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," *Presence: Teleoperators & Virtual Environments* 6, no. 6 (1997): 603–616.
[129] Roads, Curtis. "Artificial intelligence and music." *Computer Music Journal* 4.2 (1980): 13-25.
[130] Simon, Herbert A., and Richard K. Sumner. "Pattern in music." *Formal representation of human judgment. New York: Wiley* (1968): 219-250.
[131] Winograd, Terry. "Linguistics and the computer analysis of tonal harmony." *Journal of Music Theory* 12.1 (1968): 2-49.

writing.[132] In the mid 1980s, Roads proposed a roadmap for the research in music and AI, identifying problems in music composition, performance, music theory and digital sound processing. The main problem to address, in his view, is that of the 'representation' of the musical domain in the digital domain, and how do we (and the computer) make *meaning* of it. With the potential for applications in intelligent composer assistants, responsive instruments, generative modeling of music, and recognition and analysis of musical sound.[133]

1.6.2 AI music generation and Neural Audio Synthesis

Since the 1990s, this has become a field in its own right.[134] In the present day, the literature on this topic is incredibly vast, with numerous system architectures currently implemented in a variety of AI music generators.[135] When compared to digital sound synthesis there are significant differences in the process of audio generation using AI technology. One of the approaches is the use of deep neural networks for raw audio waveform modeling.[136] This means the generation of spectral content one sample at a time, in a sequential manner. There are multiple system architectures used in neural audio synthesis including: feedforward neural networks (FF), convolutional neural networks (CNN), recurrent neural networks (RNN), long-short-term memory networks (LSTM), generative adversarial networks (GANs), variational autoencoders (VAE) and transformer networks.[137] Neural networks in general have become the most widely used, with different combinations of architectures being implemented to achieve different results, such as unconditional generation of audio[138] or timbre transfer.[139] The procedure is usually computationally expensive, especially at what is considered high-quality audio sampling rates (>44.1

---

[132] Roads, Curtis. "Artificial intelligence and music." *Computer Music Journal* 4.2 (1980): 13-25.
[133] Roads, Curtis. "Research in music and artificial intelligence." *ACM Computing Surveys (CSUR)* 17.2 (1985): 163-190.
[134] Camurri, Antonio. "On the role of artificial intelligence in music research." *Journal of New Music Research* 19.2-3 (1990): 219-248.
[135] Ji, Shulei, Jing Luo, and Xinyu Yang. "A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions." *arXiv preprint arXiv:2011.06801* (2020).
[136] Van Den Oord, Aaron, et al. "Wavenet: A generative model for raw audio." *arXiv preprint arXiv:1609.03499* 12 (2016).
[137] Civit, Miguel, et al. "A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends." *Expert Systems with Applications* 209 (2022): 118190.
[138] Mehri, Soroush, et al. "SampleRNN: An unconditional end-to-end neural audio generation model." *arXiv preprint arXiv:1612.07837* (2016).
[139] Mor, Noam, et al. "A universal music translation network." *arXiv preprint arXiv:1805.07848* (2018).

kHz) and the different stages involved are usually carried out in an asynchronous manner, although there have been recent developments in real-time neural audio synthesis.[140] Multiple AI music generators are accessible today, many are products and services that one can subscribe to, and each are aimed to different audiences, with different capabilities.[141]

These developments represent a paradigm shift (maybe ontological even) compared to previous sound synthesis techniques such as additive, subtractive, FM, or granular. Neural sound synthesis is in many ways about matching prompts and inputs with the contents of a dataset to be recombined at the sample level on its output. These advancements constitute a different paradigm in the generation of sonic material when compared to digital sound synthesis. A paradigm that brings us much closer to the idea of simulation.

However, these systems are not perfect, for example, I signed up to test Google's AI kitchen,[142] an audio generator based on text prompts, when it first was released. I wanted to generate long sustained notes with a saxophone timbre, naturally the prompts would be variations of the words 'drone', 'saxophone', 'sustained notes', 'long notes', etc, but the result always had a piano accompaniment. It made me realize that most likely the 'saxophone' sounds contained in the data set must have been probably jazz recordings, where there is most likely always a piano in the background. If there are no recordings of saxophones performing long held notes, then could it not generate any? Given a very generic prompt such as 'jazz saxophone sad melody', one can expect the output to be exactly that, a muzak-esque jazz saxophone melody. Maybe I would have achieved a much closer result using other forms of synthesis. But my intention here is not to point out which synthesis technique is 'better', or more 'creative' but simply to frame the developments in these technologies as putting us in the direction of XR through ideas of simulation and replication in the sonic domain.

[140] Caillon, Antoine, and Philippe Esling. "RAVE: A variational autoencoder for fast and high-quality neural audio synthesis." *arXiv preprint arXiv:2111.05011* (2021).

[141] https://www.aiva.ai/
https://suno.com/
https://openai.com/index/musenet/

[142] https://aitestkitchen.withgoogle.com/

Chapter 1 Summary

In this chapter I have described the concepts of XR through Milgram and Kishino's taxonomy of visual displays, as an example of the heavy reliance towards the visual modality in XR. I contrast this with the emerging field of music in XR, which positions the auditory modality at the center of XR experiences.

While many of the XR technologies described focus on the visual display of virtual objects, virtual environments, and augmentations into the real world, many do implement interactions that converse with sound, as is the case of the work done by Myron Krueger and Jaron Lanier.

In an attempt to find correspondences between XR and computer music I describe landmark technological development in digital sound synthesis, spatial audio, virtual agents, and artificial intelligence. I found correspondences between the XR literature and the approaches in musical composition, and ideas from composers working with computer music. These findings provide useful terminology and helpful insights into the development of a musical practice with XR which is described in the following chapters.

Chapter 2
The XR musical landscape:
Concepts, Design Frameworks, and Creativity

This chapter consists of a survey of relevant concepts, design frameworks and creative works in music in XR. I begin by describing XR concepts such as immersion, presence, virtual embodiment and cybersickness. These concepts are necessary to understand the existing design and evaluation frameworks for music in XR applications, which I describe in the following section. I continue with a survey of the creative output of artists and scholars who work in musical XR. I finish this chapter by discussing emerging compositional trends and practices between the various examples.

2.1 XR Concepts

2.1.1 Immersion

In the context of XR, immersion can be understood in two ways: (1) as a descriptor of computer display technology and its capacity to deliver a vivid illusion of reality to the senses of a human participant. [143] [144] (2) ‘A psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an [virtual] environment that provides a continuous stream of stimuli and experiences.’[145] The first definition concerns the resolution and fidelity of multi-modal displays (visual, auditory, haptic), and the tracking capabilities of a system for proprioceptive feedback of body movements, i.e. rotating your head to look around in a virtual environment. In contrast, the second definition focuses on the experience of the user, and their perception of self in a virtual environment i.e. being isolated from the physical environment, perception of self-movement, and natural modes of interaction and control.

[143] Slater, Mel, et al. “Immersion, presence and performance in virtual environments: An experiment with tri-dimensional chess.” *Proceedings of the ACM symposium on virtual reality software and technology*. 1996.
[144] Slater, Mel, and Sylvia Wilbur. “A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments.” *Presence: Teleoperators & Virtual Environments* 6.6 (1997): 603-616.
[145] Witmer, Bob G., and Michael J. Singer. “Measuring presence in virtual environments: A presence questionnaire.” *Presence* 7.3 (1998): 225-240.

The first definition by Slater et al., treats immersion as an objective quantifiable measure that can be used to compare the degree of virtuality of systems, by describing some systems as more ‘immersive’ than others based on the technology used; for example, a VR headset is more immersive than a cinema, given that the former provides a 360º view of a virtual environment and binaural sound, plus 6DoF tracking. The cinema can only display a virtual environment in a flat 2D screen and surround sound, thus it is less immersive. Under this definition, examples of different media without 360º view, binaural audio, or tracking capabilities are not immersive. However, for Witmer and Singer, immersion is something the user experiences and thus depends on the user's perception, making it a subjective measure. They agree that using a head-mounted display (HMD) is fundamental to experiencing immersion in a virtual environment, as it isolates users from the physical environment, but that in itself is not enough when comparing how ‘immersive’ an experience is. They also consider the level of ‘involvement’ from the user, which they define as the user directing “energy and attention to a coherent set of stimuli or meaningfully related activities and events.”[146] We could think of some classic examples that would fit this definition such as reading a book, watching a movie, or playing a videogame – all of which are experiences where humans can be ‘involved’, but have low levels of immersion.

Thus, immersion can be understood as both the objective measure of the technology implemented in an XR experience and the subjective measurement of the users’ perception of a simulated environment.

2.2.2 Presence

Presence in XR is commonly described as a sense of ‘being there’, the subjective experience of being in a virtual environment, even when one is physically situated in another space.[147] This is most commonly experienced in VR, where a user is immersed in a virtual environment. It can also be extended to AR and MR as researchers in teleoperation systems found. Teleoperators would often describe a sense of being in

[146] Witmer, Bob G., and Michael J. Singer. “Measuring presence in virtual environments: A presence questionnaire.” *Presence* 7.3 (1998): 225-240.
[147] Witmer and Singer (1998).

the remote worksite rather than their control stations when performing different tasks remotely.[148] Witmer and Singer consider both involvement and immersion as necessary to experience presence. These two concepts are interdependent in their understanding of presence. They hypothesize that users' sense of presence in a virtual environment depends on their self-perception as part of the stimulus flow; as being able to affect and be affected by it, and paying attention to the continuities, connectedness and coherence of that stimulus flow.[149]

For Slater, the definition and conditions for presence are slightly different from those of Witmer and Singer. Although for Slater the actions of users immersed in virtual environments and how they perceive the virtual environment are tangled together in what he defines as sensorimotor contingencies, i.e being able to change our gaze direction and head position to look at an object from another angle in a virtual environment; for him, 'presence' only applies to immersive systems i.e. HMD or CAVE systems, making a distinction from screen-based systems and desktop computers. Instead of using the word 'presence', which has been subject to long debates spanning multiple meanings depending on the context, he prefers to break the term down to the concepts of *place illusion* (PI) and *plausibility illusion* (Psi).[150] He defines PI as the "strong illusion of being in a place in spite of the sure knowledge that you are not there." Slater hypothesizes that PI depends on two variables: (1) the match between displayed sensory data and the internal representation systems (or mental models) of a user in a virtual environment. (2) the match between proprioception and sensory data, enabling changes to the display that are consistent with changes caused by the individual's movement and locomotion.[151] The more sensory-motor contingencies that an immersive system can accurately support the stronger the PI becomes.[152] Psi is defined as the capacity for the user to believe that the actions that take place in the virtual environment are 'really'

[148] Minsky, Marvin. "Telepresence." (1980).
[149] Witmer, Bob G., and Michael J. Singer. "Measuring presence in virtual environments: A presence questionnaire." *Presence* 7.3 (1998): 225-240.
[150] Slater, Mel. "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009): 3549-3557.
[151] Slater, Mel, et al. "Immersion, presence and performance in virtual environments: An experiment with tri-dimensional chess." *Proceedings of the ACM symposium on virtual reality software and technology*. 1996.
[152] Slater, Mel, and Sylvia Wilbur. "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments." *Presence: Teleoperators & Virtual Environments* 6.6 (1997): 603-616.

happening. It can be understood as events that are not directly related to our actions in a virtual environment, but create a sensation in a participant. Psi does not necessarily need ‘physical realism’, but rather to generate realistic physiological responses in the participant. Slater uses the example of a female virtual avatar looking at the participant directly in the eye and smiling before asking a question, where the participant finds themselves smiling back and answering, although they know nobody's ‘really’ there to pick up on their smile. Such responses could be measured through physiological changes in heart rate, or skin conductivity—changes Slater associates with the ‘internal feeling provoked by an entity’ in the virtual environment.[153]

2.2.3 Virtual Embodiment

When we are presented with the illusion of being in a place by means of a HMD, it is our sense of self that is transported to this virtual environment, but what happens to the perception of our body in virtual space? The body is the vehicle through which we experience the real world. We locate our sense of self as being inside our bodies, we have a sense of ownership of our body and its different limbs – I look at my hands or legs and I know that they are my own. I can willingly control my limbs and decide what actions to undertake – ‘I will run to the bus stop’, or ‘I will grab this glass of water with my right hand’. However, discussions pertaining to the relationship between the self and the body, and how it shapes our experience of the world, have for long intrigued philosophers and cognitive scientists. Multiple experiments have been carried out by cognitive scientists to understand our embodied experience of the world and how it can possibly be manipulated. A classic experiment is the Rubber Hand Illusion,[154] where a subject is led to believe that a rubber hand is part of their body. In this experiment, the arm of the subject is hidden by a veil and presented instead with a rubber arm that resembles their own. The researchers stimulate the subject's real arm, by stroking their hidden hand with a brush, while simultaneously doing the same to the rubber hand. By means of pairing vision, touch and proprioception,

153 Slater, Mel. “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009): 3549-3557.
154 Botvinick, Matthew, and Jonathan Cohen. “Rubber hands ‘feel’ touch that eyes see.” *Nature* 391.6669 (1998): 756-756.

subjects are perceptually led to believe that what they are experiencing is the rubber hand being stroked and not their hidden hand.[155] More recently, a similar version of this experiment took place using a 3D projection of an arm instead of a rubber hand.[156]

Pairing different sensory modalities to create full body ownership illusions is a complex problem in the context of virtual reality,[157] which has direct implications and technical challenges to the development of musical performance in these environments. How to best represent my body in virtual space? A realistic or stylized virtual model to represent my body? What makes me identify this virtual body as my own? What actions can I perform with my virtual body? Are these the same as the actions I could or would carry out with my biological body? The concept of embodiment is a highly contested concept with multiple discussions across the humanities and the sciences. Researchers Kilteni et al. prefer to use the term 'Sense of Embodiment' (SoE)[158] in virtual reality. They make the distinction from the term *embodiment*, in order to move away from the more philosophical questions that derive from the definition of embodiment itself. They define the SoE toward a virtual body as the 'sense that emerges when the virtual body's properties are processed as if they were the properties of one's own biological body'. The SoE pertains to the individual experience of a person while being immersed in VR, and thus virtual embodiment plays a significant role in the design of immersive experiences, contributing to an increased sense of presence in virtual environments.[159] [160] Virtual embodiment also affects the way we perform a musical instrument in virtual reality. Researchers have observed that full body ownership illusions can change the behavior and attitude of users in the way they play a musical instrument in virtual reality depending on the appearance of the virtual body.[161] However, in the context of musical performance in

---

[155] https://www.youtube.com/watch?v=sxwn1w7MJvk, accessed December 20th, 2024, an example of the rubber hand illusion experiment in a report by the BBC.
[156] Sanchez-Vives, Maria V., et al. "Virtual hand illusion induced by visuomotor correlations." *PloS one* 5.4 (2010): e10381.
[157] Spanlang, Bernhard, et al. "How to build an embodiment lab: achieving body representation illusions in virtual reality." *Frontiers in Robotics and AI* 1 (2014): 9.
[158] Kilteni, Konstantina, Raphaela Groten, and Mel Slater. "The sense of embodiment in virtual reality." *Presence: Teleoperators and Virtual Environments* 21.4 (2012): 373-387.
[159] Slater, Mel, Bernhard Spanlang, and David Corominas. "Simulating virtual environments within virtual environments as the basis for a psychophysics of presence." *ACM transactions on graphics (TOG)* 29.4 (2010): 1-9.
[160] Biocca, Frank. "The cyborg's dilemma: Progressive embodiment in virtual environments." *Journal of computer-mediated communication* 3.2 (1997): JCMC324.
[161] Kilteni, Konstantina, Ilias Bergstrom, and Mel Slater. "Drumming in immersive virtual reality: the body shapes the way we play." *IEEE transactions on visualization and computer graphics* 19.4 (2013): 597-605.

VR it might not only be relevant for the user to have a virtual body, but for other performers and audiences that might be in the virtual environment or looking at it in a remote location through a screen. Especially in the context of music, the relation between actions and sounds, communication, and intentionality, are perceived as an embodied experience and play a fundamental role in our understanding of a musical performance.[162]

2.2.4 Cybersickness

Many users, upon entering virtual reality for the first time, have experienced feelings of vertigo, disorientation, nausea, headaches, ataxia, amongst other symptoms, in what has been labeled as *cybersickness.*[163] These symptoms closely resemble those of motion sickness. However, motion sickness and cybersickness are not the same. The former can be induced through vestibular stimulation alone, while the latter can be induced through visual stimulation.[164] There are multiple studies on cybersickness: what types of stimulus can cause it, what technological limitations contribute to it (latency, flickering), who does it affect the most in the population, and what are potential solutions to it.[165] In VR, users are usually in a stationary position in the real world and utilize joysticks for virtual navigation, relying on the illusion of self-motion or vection in a virtual environment.[166] Multiple theories have been proposed to explain cybersickness with one of the most accepted ones being that of sensory conflict theory.[167] When the body is prompted with the illusion of self-motion through visual stimulus, since actual motion is being experienced, there is a mismatch with the vestibular system. This leads the body to not know how to handle the sensory mismatch, engendering the range of symptoms that are consistent with motion sickness. However, it does not explain all of the symptoms, or why some stimuli induce cybersickness on some people while not on others.

---

[162] Corness, Greg. “The musical experience through the lens of embodiment.” *Leonardo Music Journal* 18 (2008): 21-24.
[163] LaViola Jr, Joseph J. “A discussion of cybersickness in virtual environments.” *ACM Sigchi Bulletin* 32.1 (2000): 47-56.
[164] LaViola Jr (2000).
[165] Davis, Simon, Keith Nesbitt, and Eugene Nalivaiko. “A systematic review of cybersickness.” *Proceedings of the 2014 conference on interactive entertainment*. 2014.
[166] McCauley, Michael E., and Thomas J. Sharkey. “Cybersickness: Perception of self-motion in virtual environments.” *Presence: Teleoperators & Virtual Environments* 1.3 (1992): 311-318.
[167] J. T. Reason and J. J. Brand, Motion Sickness, London: Academic press, (1975).

2.2 Design and evaluation frameworks in musical XR

In this section I review existing frameworks that outline principles and guidelines for the design and evaluation of musical instruments in XR. These frameworks cover aspects that range from technical implementation and considerations, to aesthetics, and scenographic design.

2.2.1 Designing musical instruments for virtual environments

Along with multiple examples and experiments of musical instruments and musical interactions in virtual reality, researchers have proposed different definitions, design frameworks, and evaluation dimensions. Serafin et al. outline nine design principles for the design of Virtual Reality Musical Instruments (VRMIs).[168] These principles are meant to be applied to the design and evaluation of VRMIs from the point of view of the performer. The goal is to make the experience of performing music with a VRMI smooth and efficient. They contribute relevant technical and theoretical insights on concepts that are specific to VR and digital musical instrument design, these include: avoiding latency, cybersickness, designing for multimodal feedback, leveraging existing musical skills, virtual embodiment, creating a sense of presence, making the experience social, and the implementation of 'natural' and 'magical' interactions, among other aspects.

2.2.2 Aesthetic driven design

Atherton and Wang present a philosophy of design for virtual reality that offers an interdisciplinary perspective to the design of musical VR experiences.[169] Their philosophy is centered around the balance of 'doing' versus 'being' in virtual space. They provide a set of eight lenses and associated principles to inform the design of such experiences where they pay close consideration to audio implementation, design, interaction, immersion, embodiment, playfulness and social activities. In Atherton and Wang's

[168] Stefania Serafin et al., "Virtual reality musical instruments: State of the art, design principles, and future directions," *Computer Music Journal* 40, no. 3 (2016): 22–40.
[169] Jack Atherton and Ge Wang, "Doing vs. Being: A philosophy of design for artful VR," *Journal of New Music Research* 49, no. 1 (2020): 35–59.

philosophy we find multiple themes and views that overlap with those presented by Serafin et al., including multimodality, magical or impossible interactions, virtual representations of the body and the social aspect. Other important aspects that Atherton and Wang highlight is the design in tandem of visual and sonic elements, an emphasis for design driven by aesthetics that balance action and reflection as a fundamental part of the user's experience.

2.2.3 Scenography for XR

Zappi et al. propose an evaluation system for performance configurations that use immersive and semi-immersive technologies. Their evaluation system considers the scenographic aspect of a performance in XR, how it influences the perspectives of audiences and performers, and how it could be constructed in a way that affords bi-directional interactions between the real and virtual environments.[170] The seven concepts they propose constitute a dimension space to evaluate various aspects of the audience and musicians experience through the technical choices of the performance setup. They evaluate the levels of interaction, immersion and visibility between the performers and spectators. The 'Ensemble Potential' dimension brings an interesting perspective as it evaluates the possibility of accommodating multiple performers and/or virtual instruments on stage.

2.3 Review of creative works in XR

In this section I survey the creative works of media artists, composers, and scholars who work with XR as a medium for music composition and performance. Given the multi-modal quality of XR, this review includes works that oftentimes blur the boundaries between visual arts, interactive media, digital musical instrument design, and music composition. I grouped the examples in the following categories:

- Traditional Musical Instruments
- 3D Interactions and Controllers

[170] Victor Zappi, Dario Mazzanti, and Florent Berthaut, *From the Lab to the Stage: Practical Considerations on Designing Performances with Immersive Virtual Musical Instruments*, 2022.

- Virtual Navigation
- Shared Virtual Environments
- Gamification and Video Game Mechanics
- Sandbox Environments and Audiovisual Programming
- Performances in XR

The attempt to categorize works in XR is not a simple task given its multi-modal nature, making it a fertile ground for multiple artistic streams to converge in this medium. These categories stem from finding recurrent themes in the works themselves. In implementing XR technology to develop a musical practice, these categories become helpful in painting a bigger picture of the various themes and concepts one might consider for the technical and aesthetic development of a musical composition in XR. The following figure helps visualize the relationships that exist between these categories. These relationships are not necessarily hierarchical, and they should be considered fluid, as they can govern or have a direct influence on each other. Some categories (can) contain others or be determined by the larger structures containing them. For example, in *maps and legends*,[171] the use of preexisting video game mechanics determines the 3D interactions of users in a shared virtual environment. In *Kilgore*, the use of gamification techniques and rules determines how and when users can apply certain types of 3D interactions.[172] Here, gamification can be thought of as managing a larger temporal structure of a musical composition which is carried out by the 3D interactions of the user in the virtual environment, which happen in smaller temporal units. I consider 'Virtual Navigation' as a particular dimension that traverses the multiple categories. It can be a very prominent feature of the musical composition as is the case with works like *Echo::Canyon*,[173] *Versum*,[174] and *Kilgore*,[175] or simply the means of moving around the virtual

[171] Hamilton, Robert. "maps and legends: FPS-Based Interfaces for Composition and Immersive Performance." *ICMC*. 2007.
[172] Ciciliani, Marko. "Virtual 3D environments as composition and performance spaces." *Journal of New Music Research* 49.1 (2020): 104-113
[173] Rob Hamilton, "Musical sonification of avatar physiologies, virtual flight and gesture," in *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers 10* (Springer, 2014), 518–532.
[174] Barri, Tarik. "Versum: audiovisual composing in 3d." (2009).
[175] Ciciliani (2020).

environment with no direct implication in the musical outcome.These categories are helpful to identify salient features in the design and implementation of musical works in XR, how musical form might be structured, and how these works are presented and experienced by audiences.

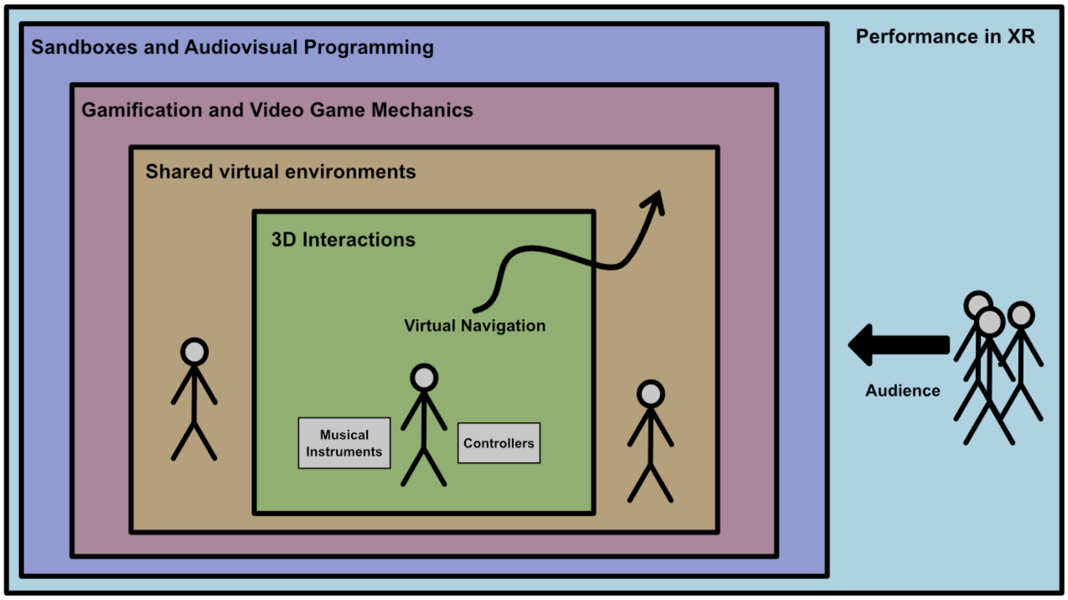


Figure 2. Relationships between categories in XR musical practice.

2.3.1 Traditional Musical Instruments

In the previous chapter, I characterized the advancements in digital sound synthesis in the 1960s as a form of virtual reality, with the virtual component deriving from the severance of sound from its mechanical mode of production. VR technology enables the integration of digital sound synthesis and a simulated mode of production. Scholars who design Virtual Reality Musical Instruments (VRMIs) recommend to leverage the expertise and knowledge required to play a real instrument onto the design of a virtual one.[176] The following examples show instruments developed in XR that take inspiration in instrumental practices with traditional musical instruments.

[176] Stefania Serafin et al., "Virtual reality musical instruments: State of the art, design principles, and future directions," *Computer Music Journal* 40, no. 3 (2016): 22–40.

2.3.1.1 Virtual Percussion

Mäki-Patola et al. developed a Virtual Xylophone, among other instruments, for a single performer in a CAVE-like environment.[177] The instrument is played by striking virtual xylophone plates by means of virtual mallets, which are tracked via magnetic sensors on each hand. Although inspired by an actual xylophone, there are properties of the virtual simulation that make it a distinct experience. For example, the performer can select different plates from a 'piano roll' interface, and position them in different points of the virtual space. The performer can run the virtual mallet through multiple plates in one single gesture without any constraints given the lack of an actual physical presence or haptic feedback. The design results in performance behaviors that are uncommon when compared to playing a real xylophone. For example, the plates can be spatially arranged in 'impossible' ways around egocentric space in order to perform a specific musical piece.

Another instrument developed with the same system is the Virtual Membrane, it is implemented with the same technology as the Virtual Xylophone and also uses virtual mallets. In this example, the complex sound models used yield different timbres as the performer hits different locations of the Virtual Membrane. They can also choose different materials and adjust the size of the membrane to impossibly large sizes. The sound is visualized as an animated wave on the virtual environment, providing audiovisual feedback of the instrument to the performer.

2.3.1.2 Virtual String Instruments

Coretet is a set of VRMIs that are inspired by the mechanics and gestures involved in the performance of traditional bowed string instruments.[178] Created by Rob Hamilton, the virtual environment supports the networked performance of up to four different virtual instruments. Inspired by the format of the traditional string quartet, the virtual quartet includes violin, viola, cello and double-bass. The explicit

---

[177] Mäki-Patola, Teemu, et al. "Experiments with virtual reality instruments." *Proceedings of the 2005 conference on New interfaces for musical expression*. 2005.

[178] Hamilton, Rob. "Coretet: a 21st century virtual interface for musical expression." *14th International Symposium on Computer Music Multidisciplinary Research*. Springer, 2019.

design goal was to develop virtual instruments that would be performed by professional string instrument musicians. The design leverages the performer's knowledge of traditional bowed string instruments to achieve a high degree of skill with the virtual instrument and potentially lead to virtuosity.[179] Coretet combines the game engine Unreal for the interactive design with the STK physical string models in Pure Data for sound synthesis. The performers share the same virtual space, where they can visually gesture cues with their heads, hands and instruments. They interact with a virtual representation of their instrument using both of their hand-held controllers. A virtual bow on the right hand can be used to excite the virtual strings of the instrument. Button inputs select which string will be activated. The left hand is used along the virtual neck to select different notes and modulate different sound parameters. The sound of the physical string model is augmented with oscillators and effects such as gain staging, reverb and compression. Using the Coretet system, the author composed the three movements of the piece *Trois Machins de la Grâce Aimante*.[180] These movements explore different performance modes: (I) improvisation, (II) traditional notated score, (III) real-time display of the score along the neck of each instrument. In the performance setting, the musicians wear HMDs and hand-held controllers while sitting on chairs arranged just like an IRL string quartet. The audience is looking at a screen where a virtual camera provides a third person view of the virtual environment with the four performers.

2.3.2 Virtual Navigation

Virtual environments afford the design of different scenarios, ranging from enclosed spaces to vast open worlds. These environments can be designed with varying degrees of constraint, limiting movement through physics (i.e gravity), determined movement speed, and the design of obstacles. The environment can also be liberated from all constraints, allowing users to 'fly' anywhere and clip through virtual

[179] Hamilton, Rob. "Coretet: a 21st century virtual interface for musical expression." *14th International Symposium on Computer Music Multidisciplinary Research*. Springer, 2019.
[180] Hamilton, Rob. "Trois Machins de la Grâce Aimante: A virtual reality string quartet." *Proceedings of the 2019 International Computer Music Conference, New York*. 2019.

objects. In this section I look into examples that use virtual navigation as a way to structure and perform musical pieces and artistic experiences.

2.3.2.1 Floating in a virtual environment

Osmose is an immersive artwork that uses ambiguity, translucency, non-linear music composition and virtual navigation to explore the virtual realm as a space that is unlike our habitual perceptions.[181] The piece is set up as both an installation and a performance. One user is totally immersed by means of a HMD. The audience, in a separate room, can perceive what the user sees and hears through the projection of the user's stereoscopic vision on a screen and a multi-channel speaker system. The user is behind a translucent screen and their silhouette can be seen by the audience. Virtual navigation in *Osmose* is controlled through a custom interface that simulates scuba diving or the experience of buoyancy.[182] Tilt sensors were used to measure the inclination of the user's spine and the expansion and contraction of their chests. The tilt controls the movement on the horizontal plane, while the breath of the user (expansion/contraction) controls the movement in the vertical plane. The interaction suggests the sensation of floating and generates a strong sense of presence for the users experiencing it.[183] The transparency of the visual elements make the virtual environment neither figurative or abstract, an effect achieved by the use of transparency to dissolve the figure and ground relationships. As the user floats around, the transition between multiple worlds are slow and subtle, with the possibility of superimposing different worlds. Along with the transition between virtual worlds and representations, sound in *Osmose* is interactive, responding to changes in the user's location, direction and speed. Navigation becomes an ambiguous experience as the user transits this non-habitual space while maintaining awareness of their balance and breath.[184]

---

[181] Davies, Char. "OSMOSE: Notes on being in Immersive virtual space." (1998): 65-74.
[182] Davies (1998).
[183] Davies, Char, and John Harrison. "Osmose: towards broadening the aesthetics of virtual reality." *ACM SIGGRAPH Computer Graphics* 30.4 (1996): 25-28.
[184] Davies (1998).

2.3.2.2 Directed navigation as performance

In *Kilgore*, an interactive piece by Marko Ciciliani, two performers navigate virtual space by controlling virtual avatars projected on a 2D screen using video game controllers on stage.[185] The performers are given different tasks that they have to complete in virtual space in order to advance the different sections of the piece. Some of these tasks involve pushing and collecting virtual objects, with performers moving their avatars through a landscape filled with obstacles that make virtual navigation a challenging effort. The position, rotation and direction of the first-person view of the virtual avatars drives different sound synthesis processes as they carry out the tasks. Throughout the piece, in specific sections, performers switch to play real instruments while the virtual avatars become automated and accompany their performance. For Ciciliani, designing the virtual space translates to composing the possibilities for musical outcome. Sequences of sound that are mapped to spatial transformations in the virtual environment are determined through the facilitation or difficulty of spatial connections.[186]

2.3.2.3 'Open world' navigation as a musical instrument

In *ECHO::Canyon* virtual navigation plays a more direct role in shaping the sonic outcome, using virtual navigation as an instrument. In this example, the music is generated from real-time sonifications of macro and micro motions of a virtual avatar.[187] The avatar is a fictional winged character that is controlled in real-time from a third person perspective using a game-pad controller or mouse and keyboard. The flight of the avatar throughout the landscape is approached as a macro gesture. By tracing rays to the sides and below it Hamilton is able to track the distance between the avatar and different features in the virtual environment and use that data to drive various sonification parameters. The data is output via OSC to SuperCollider and is mapped to different musical parameters of granular and FM synthesis, including:

[185] Ciciliani, Marko. "Virtual 3D environments as composition and performance spaces." *Journal of New Music Research* 49.1 (2020): 104-113.
[186] Ciciliani (2020).
[187] Rob Hamilton, "Musical sonification of avatar physiologies, virtual flight and gesture," in *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers 10* (Springer, 2014), 518–532.

amplitude, central frequency and grain count for the former, and amplitude, feedback and phase modulation of the latter.[188] For Hamilton, the design of the topography of the virtual environment fulfills the role of composition, creating different pathways that allow the performer to spatially interact in a diversity of ways with the virtual environment.[189]

2.3.2.4 3D space as a sequencer

Versum is a non-immersive virtual environment, controlled via a game controller, that enables the creation of audiovisual compositions.[190] Barri developed this system with the idea to subvert the usual conventions of playing sounds from left to right in sequencers and DAWs, by creating a three-dimensional space that can be used to sequence sounds. In this system, the 'actor' can fly through an abstract space populated with 'entities' that produce sound. The amplitude of the sounds emitted by the entities are dependent on the distance to the actor flying through space, with sounds getting louder as the actor gets closer to an entity. Similarly, a doppler effect is applied to the sound by calculating distance and speed, contributing to making the sounds more dynamic and interactive as it relates to the speed of the actor. The composer can rearrange the position and number of entities in the abstract space to create different audiovisual patterns.

2.3.3 3D Interactions and Controllers

XR gives designers the possibility to embed virtual objects with custom properties and interactive capabilities. Through the design of custom controllers or the augmentation of commercial ones, it is possible to create novel interactions with virtual objects to generate and manipulate sound. This opens up the possibility to generate novel VRMIs that diverge from the design and interactions found in traditional musical instruments.

---

[188] Rob Hamilton, "Musical sonification of avatar physiologies, virtual flight and gesture," in *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers 10* (Springer, 2014), 518–532.
[189] Hamilton (2014).
[190] Barri, Tarik. "Versum: audiovisual composing in 3d." (2009).

2.3.3.1 Custom physical controllers for interactions in immersive environments

Berthaut et al. have explored the creation of efficient and expressive multi-process instruments in immersive virtual environments. The DRILE virtual environment allows for the visualization and simultaneous control of the sound processes through 3D reactive widgets.[191] To control and interact with the 3D widgets the researchers developed the Piivert, a novel interface with which a user can perform different gestures by means of a 6DoF virtual ray and FSR sensors.[192] The user can point to and select a reactive widget in the virtual environment with the controller. This widget can be excited by applying pressure on the force sensors. To apply modulations to the sound of the selected widget the users can move it through a 'tunnel' that represents the values of a sound parameter or sound effect, i.e. amplitude, vibrato, reverb, among others. This tunnel can be grabbed and moved in order to affect multiple reactive widgets simultaneously. The audiovisual mappings are bidirectional, meaning that the reactive widget can display musical events while simultaneously having the graphics be affected by the sound processes. In this particular example, the immersive environment is set up with a stereoscopic screen allowing users and audience to perceive 3D objects as if they were in the real space when wearing stereoscopic glasses.

2.3.3.2 Technologies for immersive performances

In *Reflets*, musicians and audience are immersed in a mixed reality environment that allows for the display of virtual objects on the stage.[193] By combining reflective transparent panels, depth cameras and projection mapping, virtual 3D objects or topologies are projected onto the real world and can only be 'seen' or revealed when performers and audience members cut across them with their bodies, meaning that the virtual topology is projected onto them. The set up for *Reflets* allows for multiple users to interact with multiple virtual objects, maintaining visual communication and encouraging audience participation.

---

[191] Berthaut, F., Desainte-Catherine, M., Hachet, M.: "DRILE: an immersive environment for hierarchical live-looping" in Proceedings of the International Conference on New Interfaces for Musical Expression (NIME) (2010), 192–197.
[192] Berthaut, Florent, Myriam Desainte-Catherine, and Martin Hachet. "Interacting with 3D reactive widgets for musical performance." *Journal of New Music Research* 40.3 (2011): 253-263.
[193] Berthaut, Florent, et al. "Reflets: Combining and revealing spaces for musical performances." *New Interfaces for Musical Expression (NIME)*. 2015.

The set up can also be paired with real musicians and other art forms, allowing for cross interactions with virtual objects for different purposes.

2.3.3.3 Re-arranging scores in VR

In the work VR Open Scores, the authors take inspiration in a real score to create an aleatoric score-based virtual scenario. With that notion in mind and expanding on the idea of the Open Work by Eco, they created two virtual scenarios of Bruno Maderna's score 'Serenata per un satellite'.[194] The first scenario is a 360 degrees video, with the original score unfolded on a virtual sphere that surrounds the user. A recording of the score is played back, with each note spatially located in the position of the corresponding notated element. Visual cues are added to highlight the notated element as the sound plays. The second scenario has multiple fragments of the original score displayed as paintings in an art gallery setup. The user can interact with the different fragments by gazing upon them. When the raycast system intersects a fragment, it triggers its playback. If the user looks away, the fragment will play until the end of what is notated, allowing the users to overlap multiple fragments by triggering multiple displays in a short amount of time.

2.3.3.4 Magnet based haptics for VR music controllers

In this project, researchers Çamcı and Granzow combine digital fabrication and VR to develop new musical instruments that investigate sensory mappings across the visual, auditory and haptic modalities.[195] One of their 'hyperreal instruments' consists of an adapter for the Oculus Touch controller that allows for the integration of neodymium magnets.[196] The instrument is activated by bringing the controllers close to each other, with the magnets providing haptic feedback through the repelling force of the magnets. The distance and angle between the controllers is mapped to the frequency difference between two oscillators,

---

[194] Masu, Raul, et al. "VR open scores: scores as inspiration for VR scenarios," in *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME 2020* (Birmingham City University, 2020), 109–114.

[195] Çamcı, Anıl, and John Granzow. "Hyperreal instruments: Bridging VR and digital fabrication to facilitate new forms of musical expression." *Leonardo Music Journal* 29 (2019): 14-18.

[196] Çamcı, Anıl, and John Granzow. "Augmented Touch: A Mounting Adapter for Oculus Touch Controllers that Enables New Hyperreal Instruments." (2022).

and a third oscillator for amplitude modulation. The design of the adapter makes the buttons and thumbstick of the controller accessible, affording other control parameters such as recording and clearing a buffer through the buttons and the change of playback speed through the thumbstick.

2.3.4 Shared Virtual Environments

Virtual environments afford collaborative interactions between multiple users. Whether users share a virtual space locally or across a network, collaborative virtual environments provide new perspectives for the development of musical instruments in XR. Questions about virtual embodiment, communication, and music as a social activity arise in the development of shared virtual environments.

2.3.4.1 Gesture control in multiplayer music environments

Carillon is a mixed reality performance for a multiplayer collaborative virtual environment.[197] The 3D environment is a large-scale virtual carillon that produces sound by means of spinning virtual rings. Multiple performers can interact with these rings by changing their rotation speed via hand gestures that affect sound synthesis parameters. Given the large scale of the instrument in virtual space, performers interact with smaller representations of the rings that appear in their HMDs. The hand gestures of the performers are tracked with LEAP motion sensors and the sound synthesis is implemented in Pure Data, with data running from Unreal Engine to Pure Data via OSC. The performance blends pre-composed audiovisual material, in the form of bells and strikers, and sound synthesis processes that are manipulated in real-time by the performers. The audience can see the performers wearing HMDs on stage, while a virtual camera shows a third person perspective of the virtual environment.

---

[197] Hamilton, Rob, and Chris Platz. "Gesture-based collaborative virtual reality performance in carillon." *Proceedings of the 2016 international computer music conference*. 2016.

2.3.4.2 Modular synthesis in VR

Modular Reality is a system for modular synthesis in Virtual Reality or 'MSVR'.[198] The system supports networking multiple users wearing HMDs in a single virtual scene, allowing for real-time collaboration. The users can add, edit and connect different sound modules using hand-held controllers. These modules were created using Max/MSP gen objects.[199] A library of modules is available to the users, which includes most of the built-in gen operators that exist in Max/MSP plus modules that were specifically designed for MSVR. Within the virtual environment these modules can be broken to its smallest parts, VCOs, LFOs, VCAs, Sample and Hold, and can be reconnected to create new modules. The implementation of modular synths in virtual reality can overcome some of the physical limitations of modular synths IRL. For example, users can interact with modules by pointing at them from a distance or bring them closer using the joystick. They can also make infinite connections with input and output virtual jacks that can have multiple cables connected to them.

2.3.4.3 Exploring communication for music performance in shared virtual environments

In the case of collaborative music making, performers immersed in virtual environments might want to communicate through modalities other than sound, given that sound is the primary medium in this creative endeavor. To study this problem researchers Men and Bryan-Kinns designed LeMo which supports music making in a collaborative virtual environment.[200] In this virtual environment, the performers are virtually embodied through avatars (showing only a head and a pair of hands) in the same virtual space. They wear HMDs for head tracking, with LEAP motion sensors attached for hand gesture tracking. The performers have access to LeMo's 3D interface, a virtual reality step sequencer that can produce loops consisting of 8-beats. The interface consists of a 2D matrix represented in 3D space, where performers can choose what steps (columns) will play a note, from rows of musical notes (analogous to

[198] Palumbo, Michael, Alexander Zonta, and Graham Wakefield. "Modular reality: Analogues of patching in immersive space." *Journal of New Music Research* 49.1 (2020): 8-23.
[199] https://docs.cycling74.com/legacy/max8/vignettes/gen_overview accessed October 20, 2024
[200] Men, Liang, and Nick Bryan-Kinns. "LeMo: supporting collaborative music making in virtual reality." *2018 IEEE 4th VR workshop on sonic interactions for virtual environments (SIVE)*. IEEE, 2018.

most commercial step sequencers). Using this system, the authors studied how participants could communicate using 3D lines or annotations that users could trace using their hands. They found that the annotations elicited social relations such as 'drawing' between the performers, approval or disapproval of choices, managing the active and passive roles to interact with the interface, and the expression of compositional thoughts.[201]

2.3.4.4 Performing in virtual spaces

Researchers Dziwis, Coler and Porschman developed a web-based multi user virtual environment that can be accessed and experienced through a web browser on a computer or mobile device, and VR/AR head mounted displays. The environment was developed in A-Frame[202] in combination with multiple components that allow for live streaming of volumetric audio and video via Kinect, live coding in multiple languages , performance with virtual instruments developed in Pure Data, all while in a telematic/networked environment as performers and audiences experience the performance together in the virtual environment.[203]

Using this system, they performed a telematic concert in a virtual environment titled *The Entanglement*.[204] For this performance both audiences and performers were 'together' sharing an online multi-user environment. The performers were set in different locations, with audio and volumetric video being streamed into the virtual environment in real-time. The volumetric video is captured and streamed using Microsoft Kinect cameras, while the audio is spatially rendered in binaural. Audiences can move freely in the virtual environment, allowing them to experience the volumetric representation and audio spatialization from different perspectives.

201 Men, Liang, and Nick Bryan-Kinns. "LeMo: supporting collaborative music making in virtual reality." *2018 IEEE 4th VR workshop on sonic interactions for virtual environments (SIVE)*. IEEE, 2018.
202 https://aframe.io/ accessed October 23, 2024
203 Dziwis, Damian, Henrik Von Coler, and Christoph Pörschmann. "Orchestra: a toolbox for live music performances in a web-based metaverse." *Journal of the Audio Engineering Society* 71.11 (2023): 802-812.
204 Dziwis, Damian, and Henrik von Coler. "The entanglement: Volumetric music performances in a virtual metaverse environment." *Journal of Network Music and Arts* 5.1 (2023): 3.

2.3.5 Gamification and Video Game Mechanics

As exemplified in the first chapter, much of the development of virtual environments and our interactions in them have been shaped in part by the developments in the video game industry. This has contributed to researchers exploring the sonification and adaptation of already existing game mechanics for musical purposes, as well as applying gamification techniques to musical experiences.

2.3.5.1 Repurposing video game mechanics for music composition

By modifying the open-source FPS-style video game engine of Quake III,[205] Hamilton developed a system for multi-user composition for his work *maps and legends*.[206] Multiple in-game parameters such as the users' position and action data can be connected to music and sound synthesis processes and trigger pre-composed material in PureData. Different sound effects, such as reverb and chorus are modulated as the performers move close to highlighted pathways in the virtual scene, in what the author refers to as a 'compositional map'. While the virtual environment is presented on a 2D screen to the audience, the data from the virtual space is communicated via OSC to an 8-channel speaker system to spatialize the sound in physical space. The goal was to create a perceptual correspondence between the virtual and physical spaces through sonic immersion.

2.3.5.2 Game aesthetics in virtual instrument design

Fijuu2 a gamified interface that takes inspiration in the aesthetic of video games from the early 2000s.[207] It was presented as an installation/performance piece where the user can sculpt the sound of six unique 3D instruments using gamepads similar to those of the Playstation 2 gaming console. The players manipulate the 3D instruments, which are represented by multiple abstract shapes to which the player can apply

[205] Hamilton, Robert. "Building interactive networked musical environments using q3osc." *Audio Engineering Society Conference: 35th International Conference: Audio for Games*. Audio Engineering Society, 2009.
[206] Hamilton, Robert. "maps and legends: FPS-Based Interfaces for Composition and Immersive Performance." *ICMC*. 2007.
[207] Oliver, Julian, and Steven Pickles. "Fijuu2: a game-based audio-visual performance and composition engine." *Proceedings of the 7th international conference on new interfaces for musical expression*. 2007.

diverse actions such as twisting, rotating and scaling can be twisted, rotated and scaled. Users can record and loop different actions in order to layer different sounds and create interesting sonic gestures. The graphics display information about the controllers, users actions, and the different instruments by reflecting the changes produced in the sound.[208]

2.3.5.3 Custom game mechanics for music composition

new notations - for [multi] players is an interactive work made in TouchDesigner[209] by new media artist and composer Remy Siu.[210] In this system, one-to-four players can freely explore a 3D virtual space using game controllers. The virtual environment is both the score and the instrument which performers navigate. Performers interact from a first person view where their movement and actions are translated into sound. Different musical ‘instruments’ and sound synthesis processes are represented by clusters of points organized in abstract shapes that float around virtual space. These instruments are activated as performers place their camera view upon them. Performers can move closer or further away from the clusters and adjust the size of their view in order to control their interaction with the different virtual instruments. All performers exist in the same virtual space, eliciting new behaviors from the clusters when multiple players interact with them. Besides controlling the size of the rectangular selection view, each player has control over the volume of their output and can select between multiple cameras and loop the last four seconds of their view and movement. The piece challenges traditional notions of scores and traditional instruments by blending game mechanics and digital sound synthesis.

2.3.5.4 Video games as scores

Paul Turowski has developed multiple digital game-based music performances where the score is a video game played by musicians with their instruments. *SQ2* is composed for a quartet of performers that can play sustained pitches within an octave range. Performers are presented with a video game on a 2D

[208] fijuu2 demo, https://www.youtube.com/watch?v=SCeZM2yGGp4, accessed 02 December, 2024
[209] https://derivative.ca/ accessed 02 December, 2024
[210] Remy Siu, *new notations*, https://remysiu.com/new-notations-for-multi-player accessed November 26th, 2024

screen, which functions as the score that will guide them throughout the piece. The mechanics described by Turowski push the understanding that this is a video game that is played by performing specific pitches, in specific time durations, mimicking the complex button combinations required to play a video game with a game controller. [211] As the video game is driven by the sounds of the performers, an audiovisual performance is in turn articulated for the audience, allowing them to follow the musicians as they navigate the score.

2.3.6 Sandbox environments and audiovisual programming

Virtual environments and virtual reality can be used to implement novel musical instruments and structure musical compositions in different interactive forms. Some researchers and artists see the potential for developing new computer music programming paradigms that exploit the spatial and gestural capabilities of virtual interactions.

2.3.6.1 Immersive sandboxes

There have been multiple sandbox environments released as applications on the Steam VR platform that follow the paradigm of desktop digital sound synthesis environments such as Pure Data, or Max/MSP.[212] Patchworld is a commercial VR application where users can build their own musical algorithms by patching together a variety of 3D objects.[213] In this environment, users have their own virtual avatar through which they can patch together oscillators, filters, audio inputs, outputs, and more complex emulations of digital instruments such as sequencers and samplers. Users can engage in world-building activities by importing 3D models, and arranging the lighting in the scene, all of which can be scaled and positioned around the virtual environment to fit the users vision. These non-musical elements (3D objects

---

[211] Turowski, Paul. “SQ2.” (2019). https://livrepository.liverpool.ac.uk/3104992/

[212] Andersson, Nikolaj, Cumhur Erkut, and Stefania Serafin. “Immersive audio programming in a virtual reality sandbox.” *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society, 2019.

[213] https://patchxr.com/ accessed December 13, 2024

and lighting) can also be mapped to be controlled or reactive to sound, making it quite a versatile tool to develop aesthetic immersive performances.

2.3.6.2 Audiovisual programming

ChuGL is a text based programming language, based on ChucK, that focuses on the integration of audio synthesis and graphics rendering tightly timed synchronization in a single programming environment.[214] The disparity between audio synthesis and graphics rendering has led artists and researchers to rely on the combination of different softwares (and hardwares) to create audiovisual compositions and experiences i.e Unity, Unreal Engine for graphics, Pure Data, Max/MSP, SuperCollider for audio. The goal of ChuGL is to unify audio and graphics, providing developers with a workflow that allows them to think in audiovisual terms without having to jump back and forth between programs and paradigms for audio and video.[215]

2.3.6.3 Multisensory programming

GENESIS-RT is a full multisensory VR platform for musical creation developed by Florens et al., it uses a modular environment for the construction of virtual instruments and sound objects based on physical modelling techniques[216] that are paired with a high-quality haptics simulation system.[217] Using a virtual interface, researchers can create physical models to simulate the sound of bells, percussions, and bowed instruments and to explore physical models that do not exist in the real world. These models are coupled with different haptic interfaces to develop visual, haptic, and auditory simulations of novel musical instruments.

---

[214] Zhu, Andrew, and Ge Wang. "ChuGL: Unified Audiovisual Programming in ChucK." *Proceedings of the International Conference on New Interfaces for Musical Expression*. 2024.
[215] Zhu and Wang (2024).
[216] Leonard, James, et al. "A virtual reality platform for musical creation: GENESIS-RT." *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers 10*. Springer International Publishing, 2014.
[217] Florens, Jean-Loup, et al. "ERGOS: Multi-degrees of freedom and versatile force-feedback panoply." *EuroHaptics 2004*. 2004.

2.3.7 Live performances in XR

Through mixed reality technology artists attempt to balance the experience of performers and audience in performances involving virtual environments. In many of the works we have looked at so far, there is a disparity in the degree of immersion experienced by performers and audiences, with unintended consequences for the perception of a creative work or performance.

2.3.7.1 Multi-user mixed reality

Researchers Zappi et al. have introduced a multimodal platform for hybrid reality live performances.[218] The platform combines 3D stereoscopic projection and motion capture to overlap interactions between the real world and virtual reality by having them share and overlap on the performance stage. The characteristic of this platform is that performers and audience members can manipulate sound parameters as is the case with the performance of *VIRTUAL_REAL*.[219] In this performance the authors collaborated with an electronic music artist in the creation of custom 3D environments and 3D objects to accompany, and react to the music composition. The audience members wore stereoscopic goggles, which allowed them to see the 3D virtual objects immersed together in the same space. Through hand tracking of the spectators' hands audiences can interact with the 3D virtual objects to modulate sound. Attendees could interact with the virtual objects but they were limited to manipulating characteristics such as color or shape. The authors have conceptualized as 'Hybrid Choreography' these types of interactions with virtual graphic environments through a set of rules that define meaningful interactions in a mixed reality platform.[220]

---

[218] Zappi, Victor, et al. "Design and Evaluation of a Hybrid Reality Performance." *NIME*. Vol. 11. 2011.
[219] Zapp et al. (2011).
[220] Zapp et al. (2011).

2.3.7.2 VR conductor

Resilience is a work by Jack Atherton that incorporates a VR performer as a conductor in a laptop orchestra.[221] While the conductor wears a HMD, their view is projected onto a screen for both the audience and the ensemble to see. The piece is based on the 'aesthetic exploration of the emotional life-cycle of a plant', going through the different stages of this process.[222] Each of these stages are represented in a virtual space and involve different hand movements and gestures from the conductor. While performing these gestures, the conductor interacts with the virtual environment, advancing the structure of the musical piece in virtual space. As this happens, the laptop ensemble follows the gestures of the conductor in order to produce different sounds. The view of the virtual space is managed by the conductor, but the interactions from the laptop ensemble also have consequences and effects on the virtual environment, creating a synergy between the real world gestures and virtual reality in what can be categorized as a mixed reality performance with a laptop ensemble.

2.3.7.3 Augmenting reality through sound

Proprius is an augmented reality music composition based on the sonification of artificial animal behavior.[223] The listener wears headphones while their movements are tracked in a designated physical space via a Microsoft Kinect camera. The physical space is sonically augmented by *Propius*, allowing the listener to perceive the sounds of artificial animals, whose positioning is modeled after behaviors such as fleeing, wandering, pursuing, amongst others. The sounds produced by each agent are spatialized and rendered into a binaural scene as the listener moves through the physical space. The spatialization implementation considers both the positions of the listener and virtual agents to render the binaural audio scene in real-time. The composition is structured in five movements, where on each one a new layer of

---

[221] Atherton, Jack, and Ge Wang. "Curating Perspectives: Incorporating Virtual Reality into Laptop Orchestra Performance." *NIME*. 2020.
[222] Atherton and Wang (2020).
[223] Özcan, Zeynep, and Anil Çamci. "An augmented reality music composition based on the sonification of animal behavior." *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2018.

animals are added and superimposed. The introduction of new layers elicits different behaviours from the artificial agents, which are based on the relationships between trophic structures in an ecosystem.

2.3.7.4 Sampling sound, indexing space

Gregory Beller developed a series of mixed reality improvisation performances titled *Air Sampling*. In these performances, Beller uses a Meta Quest 2 headset and controllers to sample and playback sounds captured from a live musician using spatial interactions.[224] The samples are distributed in 3D space as the performer records both audio and gesture in real time. The 'Spatial Sampler' and 'Sound Space' interaction modes allow for different ways of mapping sound and space.[225] In the first one, the performer positions a hand in space, presses a button to start recording and utters a sound. When the performer releases the button to stop the recording, the sample is indexed to the position of the hand in that moment. The performer can 'touch' that point to play the associated recording, emulating most commercial samplers with a spatial twist. In the second mode of interaction, sound and gesture are recorded and indexed as *xyz* points, associating hand position with a timestamp of the audio recording. Beller has carried out multiple performances with this system in collaboration with live musicians.[226]

2.4 Emerging Trends in Composition in Musical XR

In this section I identify and discuss compositional trends regarding musical form, spatial interactions, action-sound relationships, and performance setups across the works surveyed previously. These discussions will serve as further discussion points for the following chapters regarding my own creative work in musical XR.

---

[224] Beller, Grégory. "Spatial Sampling in Mixed Reality." *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2023.
[225] Beller, Greg. "Sound space and spatial sampler." *Proceedings of the 2nd International Workshop on Movement and Computing*. 2015.
[226] Air Sampling #004, https://www.youtube.com/watch?v=AMq8mJEIbYw accessed December 16, 2024

2.4.1 Musical form

Throughout the development of digital musical instruments (DMIs) and compositions with interactive media, many have argued for the development of musical pedagogies for interactive instruments,[227] and musical methods for new technologies,[228] in order to develop longevity in certain practices.[229] Engaging in musical activities such as composition, performance and improvisation through the design and implementation of interactive technologies follows these lines of inquiry, particularly in the development of musical form. In this section I argue how the musical forms in many of the reviewed examples inhabit the definitions of meta-composition[230] and composed instruments.[231]

In both *Echo::Canyon*[232] and *Versum*,[233] virtual space is used to determine the form of the musical piece, but also the interactive and performative aspects of the virtual environment. In both examples, the piece is performed by moving through virtual space. In *Versum* the performer moves around to hear the different sound objects (or 'entities') that have been laid out in an open and obstacle-less virtual space. The distances and grouping of multiple entities become by analogy the notes on a scoresheet, or the possible notes to be played. It is a non-linear score, however, as it is determined by the possibilities of movement afforded to the performer, allowing for a different interpretation of a single layout of sonic entities. In contrast, in *Echo::Canyon* it is the position and speed of the performer's virtual avatar what determines the sonic outcome of the system. The design of the virtual landscape, the valleys and hills (together with a few sound objects) act as obstacles that force and constrain the movement of the virtual avatar, thus determining the possibilities of the sonic outcome.

The concept of meta-composition was developed by Curtis Bahn, and it provides a suitable

---

[227] Butler, Jennifer. "Creating Pedagogical Etudes for Interactive Instruments." *NIME*. 2008.
[228] Zbyszynski, Michael. "An Elementary Method for Tablet." *NIME*. 2008.
[229] Marquez-Borbon, Adnan, and Juan Pablo Martinez Avila. "The problem of DMI adoption and longevity: Envisioning a NIME performance pedagogy." (2018).
[230] Bahn, Curtis Robert. *Composition, improvisation and meta-composition*. Princeton University, 1997.
[231] Schnell, Norbert, and Marc Battier. "Introducing composed instruments, technical and musicological implications." *Proceedings of the 2002 conference on New interfaces for musical expression*. 2002.
[232] Rob Hamilton, "Musical sonification of avatar physiologies, virtual flight and gesture," in *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers 10* (Springer, 2014), 518–532.
[233] Barri, Tarik. "Versum: audiovisual composing in 3d." (2009).

framework for understanding the musical form of these two examples, and how form is shaped by their design:

> "A 'meta-composition' is a composition that itself composes, or facilitates composition/performance. It can be a construct of media, oral/aural transmission, and/or electronic technology. The meta-composition informs conventional musical activities, such as composition, improvisation or performance, yet itself does not prescribe a specific time based musical entity. It can range from a set of rules for improvisation, to an extensive interactive multi-media computer interface which dynamically coordinates musical information in many forms. As with composition, meta-composition is both an abstraction of a musical idea and an activity; a noun and a verb" [234]

The two systems inform composition, improvisation and performance with many similarities and differences in how they coordinate musical information given the rules and metaphors that govern the user interactions within these systems. Similarly, the work of Turowski explores the use of rule-based systems in the form of video game mechanics driven by music. The performers are given specific tasks that they complete by playing different pitches, durations and gestures in order to advance the 2D video game.[235]

But let us look at a different example in Coretet.[236] In this system, we are working with a simulation or virtualization of a string quartet. The composition of *Trois Machins de la Grâce Aimante*[237] is for performers using this system. The composition explores the use of different methods for scoring within Coretet. This means that the composer could write distinct and new pieces of prescribed time durations using the same scoring mechanisms within the virtual environment and yield new sonic

[234] Bahn, Curtis Robert. *Composition, improvisation and meta-composition*. Princeton University, 1997.
[235] Turowski, Paul. "SQ2." (2019). https://livrepository.liverpool.ac.uk/3104992/
[236] Hamilton, Rob. "Coretet: a 21st century virtual interface for musical expression." *14th International Symposium on Computer Music Multidisciplinary Research*. Springer, 2019.
[237] Hamilton, Rob. "Trois Machins de la Grâce Aimante: A virtual reality string quartet." *Proceedings of the 2019 International Computer Music Conference, New York*. 2019.

outcomes. In this sense, Coretet as a system, is in itself a meta-composition, but it could also be understood under the notion of a ‘composed instrument’ by Schnell and Battier where:

> “The term of the composed instrument underlines the fact that computer systems used in musical performance carry as much the notion of an instrument as that of a score, in the sense of determining various aspects of a musical work. (...) As a musical instrument, it should enable the performer enough degrees of liberty to explore personal and original ways of playing with it. As a machine, it is under the control of complex computational and algorithmic layers. The representation integrates the two first categories. Composers use the representational nature of the system to define events, write scores and specify the computational and algorithmic layers while performers can apply gestural controls and adjust parameters” [238]

Other gamified approaches also fit the definition of composed instrument. In the case of *Kilgore,*[239] performers navigate around virtual space seeking to complete a task as part of the score; these gestures (virtual navigation) generate streams of data that indirectly affect the piece on a sonic level.[240] In *Maps & Legends*[241] performers are suggested to follow the ‘compositional map’ proposed by the composer, while the sounds are governed by the mechanics of the game. This resonates with other works such as *new notations - for [multi] players*,[242] where the representation is the score itself, with virtual navigation being utilized as a way to advance and play the score simultaneously. The composed instrument or in these cases, the composed virtual environments balance the agency of the performers with the indeterminacy inherent in the design of the system. Composing the methods for virtual navigation, together with the game mechanics is how the representational nature of the system notates how, when and where musical

238 Schnell, Norbert, and Marc Battier. “Introducing composed instruments, technical and musicological implications.” *Proceedings of the 2002 conference on New interfaces for musical expression*. 2002.
239 Ciciliani, Marko. “Virtual 3D environments as composition and performance spaces.” *Journal of New Music Research* 49.1 (2020): 104-113.
240 Ciciliani (2020).
241 Hamilton, Robert. “maps and legends: FPS-Based Interfaces for Composition and Immersive Performance.” *ICMC*. 2007.
242 Remy Siu, *new notations*, https://remysiu.com/new-notations-for-multi-player accessed November 26th, 2024,

events will happen throughout the piece, albeit many times being non-deterministic in their larger structure. In many cases these types of systems could be interpreted as meta-compositions or composed instruments, since they yield structures that are non-linear and are not tied to a determined time entity.

In *Resilience*,[243] we can find a challenge to both the notions of meta-composition and composed instrument, by applying a linear development to the piece in the form of a narrative arc. The performers and audience are carried through the narrative by representing the life cycle of plants in the virtual environment. The narrative arc serves as a defined path that is traversed by the conductor, while the ensemble follows their gestures in order to generate the music that accompanies this journey. Although a second performance of this piece might yield different sonic results, the structure of the piece is defined and repeatable. The 'movements' or 'stages' of the piece remain the same and are accessed in a linear fashion. This highlights the possibilities of virtual environments to entertain ideas and structures related to narration and storytelling through audiovisual representations.

Although the works mentioned above are quite diverse in the technologies implemented – the interaction design, the sonic mappings, and the experience of the performers and audience – what these examples foreground is the idea of virtual space as something to be composed, whether as a score for IRL musicians, or as an XR system for musical performance.

2.4.2 Performing music through spatial interactions in XR

One of the main advantages of current commercially available VR/AR systems is their tracking capabilities, which allow for 6DoF movement through different methods. The most common method is inside-out tracking, which does not require any external equipment besides the headset itself, with newer headsets including capabilities for hand and eye tracking.[244] This makes it simple for artists to quickly prototype musical interactions using hand and head movement in VR. Here I look into the spatial metaphors used in some of the works reviewed. These design choices present affordances and constraints

---

[243] Atherton, Jack, and Ge Wang. "Curating Perspectives: Incorporating Virtual Reality into Laptop Orchestra Performance." *NIME*. 2020.
[244] Alanko, Suvi. "Comparing Inside-out and Outside-in Tracking in Virtual Reality." (2023).

for the development of musical gestures and their articulation into larger musical forms.

There has been thorough research on the relationship between sound, movement, and meaning,[245] which has been echoed in the development of movement and gesture based interactions for musical performance and composition.[246] Since the middle of the 2000s, there have been multiple examples of DMIs that have been developed with IMUs,[247] gesture recognition,[248] and motion tracking.[249] In the field of musical XR, Berthaut et al. have researched and categorized the wide range of 3D interactions that can be used for musical expression.[250] This includes interactions with 3DUIs like buttons, sliders, 2D and 3D graphs; and 3D interaction techniques including selection, manipulation, raycasting and forms of virtual navigation.[251] XR adds another layer for musical interaction by creating a virtual world reference. Whether in VR or AR, space becomes both a container of information and a space for action, where movement is performed in relation to the virtual environment and the virtual objects in it. XR expands the relationship between movement and sound through the design of virtual object-action-object systems that can create novel action-sound relationships.[252] Authors Deacon and Barthet have analyzed and categorized the design of spatial interactions of several VR applications for interactive audio systems. They highlight the various roles of space in musical experiences in VR, where it can be a holder of interactive elements, the medium for the sonic experience, or act as a visual resource for enhancing musical performance.[253]

The Virtual Xylophone[254] is a good example where the uncommon spatial arrangements engender uncommon performance behaviors when compared to a real xylophone, as performers can position and

---

[245] Godøy, Rolf Inge, and Marc Leman, eds. *Musical gestures: Sound, movement, and meaning*. Routledge, 2010.
[246] Lewis, George E. "The virtual discourses of Pamela Z." *Journal of the Society for American Music* 1.1 (2007): 57-77.
[247] Medeiros, Carolina Brum, and Marcelo M. Wanderley. "A comprehensive review of sensors and instrumentation methods in devices for musical expression." *Sensors* 14.8 (2014): 13556-13591.
[248] Visi, Federico, Rodrigo Schramm, and Eduardo Miranda. "Gesture in performance with traditional musical instruments and electronics: Use of embodied music cognition and multimodal motion capture to design gestural mapping strategies." *Proceedings of the 2014 International Workshop on Movement and Computing*. 2014.
[249] Miranda, Eduardo Reck, and Marcelo M. Wanderley. *New digital musical instruments: control and interaction beyond the keyboard*. Vol. 21. AR Editions, Inc., 2006.
[250] Berthaut, Florent. "3D interaction techniques for musical expression." *Journal of New Music Research* 49.1 (2020): 60-72.
[251] Berthaut (2020).
[252] Jensenius, Alexander Refsum. "Action-sound: Developing methods and tools to study music-related body movement." (2007).
[253] Deacon, Thomas, and Mathieu Barthet. "Spatial Design Considerations for Interactive Audio in Virtual Reality." *Sonic Interactions in Virtual Environments*. Cham: Springer International Publishing, 2022. 181-217.
[254] Mäki-Patola, Teemu, et al. "Experiments with virtual reality instruments." *Proceedings of the 2005 conference on New interfaces for musical expression*. 2005.

stack xylophone bars in impossible positions, but always in an egocentric arrangement (and limited by the physical space). The Coretet system[255] also relies on egocentric space as the virtual string instruments are projected onto the hands of the performers and they move their controllers in order to activate and play their instruments using only motion to perform musical gestures. Similarly, in *Carilllon*[256] performers are able to interact with a large-scale instrument via hand tracking. With their hand gestures being mapped and passed on to the different control parameters that drive sound synthesis.

Although the examples categorized under 'Virtual Navigation' use screen-based configurations and game controllers – with no tracking of human bodies nor immersion – spatial metaphors play an immense role in these compositions. Ciciliani points out how the topological design of a 3D virtual environment can shape a composition's time factor, contriving the movement of virtual avatars and affecting the possibilities for performers to articulate rhythmic and melodic sequences as well as the formal structure of a piece.[257]

In the case of immersive systems, Greg Beller's Spatial Mixer[258] is a great example that exploits the spatial capabilities of XR for music interaction and performance. He explores spatial interactions by generating 3D topologies made of sound slices. In this case, sound is spatialized, but not as we usually understand it, i.e. as sources playing from different locations through a multi-channel speaker system. In this XR system, the performer moves their controllers as sound is sampled in thin discs (like ham slices) and distributed along the performer's arm movement. This creates an interactive 3D topology around the egocentric space of action of the performer. After recording both hand movement and sound simultaneously, the performer can move their hand through the aligned discs to play a sound in reverse, or at different speeds. The hand movement is indirectly linked to sound through the interaction with the virtual slices. This contrasts with the way space is used in earlier electronic instruments like the

[255] Hamilton, Rob. "Coretet: a 21st century virtual interface for musical expression." *14th International Symposium on Computer Music Multidisciplinary Research*. Springer, 2019.
[256] Hamilton, Rob, and Chris Platz. "Gesture-based collaborative virtual reality performance in carillon." *Proceedings of the 2016 international computer music conference*. 2016.
[257] Ciciliani, Marko. "Virtual 3D environments as composition and performance spaces." *Journal of New Music Research* 49.1 (2020): 104-113.
[258] Beller, Grégory. "Spatial Sampling in Mixed Reality." *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2023.

Theremin, where pitch and volume parameters are continuously related to the distance between the hands and the two antennas. In Beller's Spatial Mixer, sound is turned into discrete spatial quanta, distributed in space in accordance with the performer's hand movements at the moment of recording. The performer can use space to contain musical information and also as a means of performing that same information.

In the examples I have discussed, interactions based on movement and space are a prominent feature of multiple musical applications and performances with XR systems. These works highlight the possibilities for touchless interfaces and the use of virtual space for music composition, foregrounding spatiality as the main feature of many music interactive systems in XR.

2.4.3 Virtual re-embodiment of sound

Different sound production techniques are used throughout all of the works, including different forms of synthesis, processing of live-input, pre-recorded material, and looping. Going back to the idea in my first chapter of disembodied sound, whether it is recording or synthesis, in the context of XR sound finds a way to be virtually re-embodied. In theory, any recorded sound or sound synthesis process can be mapped and linked to any virtual interaction in XR. We could categorize all of the interactions in virtual spaces as action-sound relationships. Action-sound relationships are not natural (i.e. not coupled to a mechanical means of production), they are electronically designed.[259] In this type of relationship, Jensenius makes the distinction between electronic devices and virtual realities. The former includes door bells, mobile phones and musical instruments (DMIs), where the sound is a result from a physical interaction between the user and some mechanical part of the device, e.g. pressing a button. While the latter includes TV, movies, computer software and computer games; where the action-sound relationships are based on virtual object-action-object systems, even though these virtual systems may be controlled with a physical controller.[260]

This can be done in the fashion of looking for a sense of realism (or simulation), as in the

[259] Jensenius, Alexander Refsum. "Action-sound: Developing methods and tools to study music-related body movement." (2007), page 28.
[260] Jensenius, (2007), page 28.

examples of the Virtual Xylophone[261] and Coretet,[262] where both use physical modelling to synthesize the simulated physical properties of percussive and string sounds respectively. It seems natural to map this type of sound synthesis, as other forms might yield the sense of breaking what Slater calls 'plausibility illusion' and 'place illusion', both which create a sense of presence. An inappropriate sound choice would reveal the irreality of the instrument. Imagine playing the Virtual Xylophone and having the small virtual plates sound like timpani, we would not easily accept it as 'realistic'. But to what extent do things need to be realistic in XR? Although the Virtual Xylophone is not physically real, our expectations of sound might still be to hear something similar to a xylophone. But, what is the range of sounds that would be acceptable for us to suspend our disbelief? When does it cross that threshold and break the illusion?

Other examples seem to lend themselves more easily to abstract and arbitrary sound design, or mapping choices. For instance in *Versum*,[263] the sounds emitted by the 'entities' in the system consist of drone-like synths where the position from the performer in virtual space determines the amplitude of the sounds. The 'entities' in the system look like glowing stars, with their visualization neatly mapped to the sonic parameters as well. It is not a stretch to imagine that multiple different sound designs could have been applied to this interactive system. In this case sound synthesis, a sound making process that in its origin is already divorced from a mechanical mode of production, lends itself to be re-embodied in an abstract figure as the sonic 'entities' are in this case. String sounds or any other sustained instrument, might be too recognizable and could be perceived as inconsistent with what is being seen. Another aspect that helps tie together the sound to the 'entities' is the use of simple spatialization (stereo) and doppler effect to accentuate the relationships between speed and movement in virtual space in relation to the 'entities'.

Carillon[264] is a piece in which the sound design and mapping might sit somewhere in the middle

[261] Mäki-Patola, Teemu, et al. "Experiments with virtual reality instruments." *Proceedings of the 2005 conference on New interfaces for musical expression*. 2005.
[262] Hamilton, Rob. "Coretet: a 21st century virtual interface for musical expression." *14th International Symposium on Computer Music Multidisciplinary Research*. Springer, 2019.
[263] Barri, Tarik. "Versum: audiovisual composing in 3d." (2009).
[264] Hamilton, Rob, and Chris Platz. "Gesture-based collaborative virtual reality performance in carillon." *Proceedings of the 2016 international computer music conference*. 2016.

between the two previous examples. In this piece, performers interact with a large-scale virtual carillon using hand gestures. The sounds in the piece are a mixture of ambient sounds together with identifiable bell sounds (although synthesized). The piece traverses the sonic space between creating a sense of realism, where an audience can identify the bell sounds and associate them with a simulated mode of production. In this piece, performers alter the speed of 'rings' that drive the synthesis of bell sounds, an interaction that is nowhere near the reality of bell sounds which are percussed by mallets.

The virtual environment in *new notations - for [multi] players*[265] is completely abstract, a blank black space filled with cluster figures made of white lines and dots. However, every action is sonified in a style reminiscent of video games. Moving around is sonified as if moving through a dotted line, adjusting the bounding box in the player has a sound reminiscent of focusing a camera lens figures every action is sonified. The sounds are clearly differentiated between playing the 'instruments' in the space, and the sounds that convey to the performer how they are moving, changing their camera view and engaging in more complex interactions as live-looping the sounds tied to what the camera is focusing on.

In *Resilience* the gestures from the laptop ensemble become virtually embodied through multimodal representations. Atherton gives the example of the consistency between the gestures of performers controlling the 'wind' in the virtual environment, while also using the 'ahh' timbre to represent it sonically achieved by pairing the gestures in the virtual environment.[266] Here we find gestures that find consistency across the real world and its virtual counterpart in both sonic and visual modalities. This cross-modality between the real and virtual worlds through gesture is also expressed in the work with 'hyperreal instruments' by Çamcı and Granzow, where the micro-motions by the performer face pushback from the magnets attached to controllers creating a sense of effort in performing with this instrument.[267]

Many of the examples shown, move somewhere along this axis where sound design can be

[265] Remy Siu, *new notations*, Accessed on November 26th, 2024, https://remysiu.com/new-notations-for-multi-player.
[266] Atherton, Jack, and Ge Wang. "Curating Perspectives: Incorporating Virtual Reality into Laptop Orchestra Performance." *NIME*. 2020.
[267] Çamcı, Anıl, and John Granzow. "Augmented Touch: A Mounting Adapter for Oculus Touch Controllers that Enables New Hyperreal Instruments." (2022).

‘realistic’ or ‘abstract’, which in turn is highly dependent and intersects with many of the design principles proposed by researchers: design for multimodal feedback, interactivity, aesthetics. As pointed out previously, sound plays a large role in how we perceive a virtual environment, and the more so if we are immersed in it. And the parameters not only include sound design, but the use of spatialization, and complex mappings that would yield ‘realistic’ or ‘believable’ behaviors.

2.4.4 Experiencing XR music compositions

Throughout the many examples reviewed we find a wide range of differences in how these systems are presented to audiences in the context of musical performance. Many implement 2D screen settings where both audiences and performers perceive the interactions in 3D environments from the same perspective, as is the case with *Versum*,[268] *Fijuu2*,[269] and *new notations - for [multi] players*.[270] Other examples also use screen-based settings but implement an immersive sound system for audiences using multi-channel speaker arrays, as is the case in *Echo::Canyon*[271] and *Maps & Legends*.[272] Hamilton’s idea in these two examples was to superimpose the sonic spatialization of the virtual environment with the real world for a more immersive experience from the audience perspective. *Kilgore* is also presented on a screen for the audience, with the performers on stage switching between electronic and electroacoustic instruments to game controllers, transparenting the back and forth between the real world and the virtual environment, mediated through the performers.

Then there are examples where at least one of the performers is immersed in a virtual environment while their view is presented to the audience on a 2D screen. In *Trois Machins de la Grâce Aimante*,[273] all four performers are immersed in VR, while the audience sees both their gestures in the real

[268] Barri, Tarik. “Versum: audiovisual composing in 3d.” (2009).
[269] Oliver, Julian, and Steven Pickles. “Fijuu2: a game-based audio-visual performance and composition engine.” *Proceedings of the 7th international conference on new interfaces for musical expression*. 2007.
[270] Remy Siu, *new notations* https://remysiu.com/new-notations-for-multi-player accessed November 26th, 2024
[271] Hamilton, Rob “Musical sonification of avatar physiologies, virtual flight and gesture,” in *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers 10* (Springer, 2014), 518–532.
[272] Hamilton, Robert. “maps and legends: FPS-Based Interfaces for Composition and Immersive Performance.” *ICMC*. 2007.
[273] Hamilton, Rob. “Trois Machins de la Grâce Aimante: A virtual reality string quartet.” *Proceedings of the 2019 International Computer Music Conference, New York*. 2019.

world and a third person view of the whole virtual string quartet projected on the screen, rather than a single perspective from one of the performers. In the improvisation series with the Spatial Mixer, Beller is immersed while interacting with a musician IRL, while his view is projected onto a wall. In *Resilience*,[274] the conductor is immersed in VR, while remaining visible for the ensemble and audience to follow their gestures as well as the virtual environment on the screen. *Osmose*[275] presents a particular approach that combines multiple modalities and exploits the constraints of HMDs with aesthetic inventiveness. The piece is presented as both an installation and a performance, as it requires someone to wear a HMD to interact with the virtual environment. Meanwhile, the audience can see the first-person view of the performer on a 2D screen, while being sonically immersed in a multi-channel sound system. The performer can also be seen behind a veil, showing only their silhouette to the audience as they float in virtual space. This example solves the constraints presented by XR technology at the time by configuring the gallery space as a multi-modal performance environment where audiences can perceive the multiple levels in which the performer is executing actions.

Other examples maintain audiences and performers at the same level of immersion, as is the case with performances of DRILE[276] where both performers and audiences can see the 3D space projects on the wall, as the performer uses a custom controller to interact with the virtual environment. In *VIRTUAL_REAL*[277] both the performer and audience share a mixed reality configuration as the 3D objects are projected onto the shared physical space using stereoscopic techniques. In *Reflets* the virtual objects that are used to control sound effects are projected onto musicians IRL, superimposing the virtual space onto a real world stage where audiences can see the musicians interact with their instruments and adjust parameters to manipulate sound through the projected 3D objects. Another example where performers and audience share a similar degree of immersion is presented in *The Entanglement*,[278] where IRL musicians

[274] Atherton, Jack, and Ge Wang. "Curating Perspectives: Incorporating Virtual Reality into Laptop Orchestra Performance." *NIME*. 2020.
[275] Davies, Char. "OSMOSE: Notes on being in Immersive virtual space." (1998): 65-74.
[276] Berthaut, F., Desainte-Catherine, M., Hachet, M.: DRILE: an immersive environment for hierarchical live-looping in Proceedings of the International Conference on New Interfaces for Musical Expression (NIME) (2010), 192–197.
[277] Zappi, Victor, et al. "Design and Evaluation of a Hybrid Reality Performance." *NIME*. Vol. 11. 2011.
[278] Dziwis, Damian, Henrik Von Coler, and Christoph Pörschmann. "Orchestra: a toolbox for live music performances in a web-based metaverse." *Journal of the Audio Engineering Society* 71.11 (2023): 802-812.

are presented in a virtual environment via volumetric capture while audiences attend the virtual environment via desktop computers or HMDs. The performance takes place in the virtual environment, where audiences are allowed to roam free and explore the performance space, placing them on equal footing with the performers.

These examples show the range of different configurations used by artists to present work that involves virtual environments, HMDs and other forms of 3D interactions in virtual spaces. My intent is not to comment on which is a 'better' or 'worse' form to present these types of works. However, it is not only a problem of technical limitations, but also a problem of aesthetic implementation, and how ideas of immersion, or the different perspectives experienced by performers and audiences can be articulated in a meaningful way.

Chapter 2 Summary

In this chapter I have reviewed concepts, design frameworks and surveyed creative works that constitute the musical landscape in XR. I have identified and discussed emerging trends in the compositional and performance practices with XR systems including: the expression of musical form, the use of spatial interactions as a prominent feature for composition and interaction design, the various approaches in developing action-sound relationships to re-embody sound in virtual objects within virtual spaces, and how these performances are experienced by audiences.

Chapter 3
Music composition and performance along the virtuality continuum

In this chapter I describe the different tools, interactive music systems and creative works that I have developed in the past five years using different XR technologies. I describe the technological implementation and creative outcome for each of these works, and contextualize them by discussing related research.

3.1 VR-Mapper: a Max For Live device for prototyping movement-based sound interactions in virtual reality

I developed VR-Mapper as a response to the constraints presented by the COVID-19 pandemic in 2020. The project stemmed from my interest in developing movement-based musical interactions using motion capture technology. However, due to the lockdown of university facilities my access to the Media Interaction Lab at UVA was restricted. Using a commercial Head Mounted Display (HMD) and hand-held controllers, its tracking capabilities could serve to develop movement-based sound interactions. The result was a Max For Live (M4L) device that can receive Open Sound Control data from the Unity game development engine and map it to any parameter available on tracks and audio effects in Ableton Live.

3.1.1 Touchless interfaces and movement-based interactions

Multiple researchers and artists have implemented different technologies to develop interfaces for movement-based musical interactions. These technologies range from touchless interfaces through video tracking technology, to marker-based motion capture technology, and wearable devices with integrated sensors.

In the 1980s David Rokeby started developing his 'Very Nervous System' by using video cameras to monitor the user's action and respond with synthesized sounds in relation to the analysis of the movements captured by the camera. The cameras cover a large portion of space making the interface ‘invisible and diffuse’, with initial interactions that are not very clear, but that evolve as the audience explores and experiences the interface.[279]

Sound artist Laetitia Sonami created the *Lady’s glove,* a wearable musical instrument that has a vast array of integrated sensors which she uses to perform music. The design has gone through multiple iterations and has included different types of sensors: Hall effect, resistive strips, pressure pads, ultrasonic, and accelerometers. The data is used to control different mappings in Max/MSP which vary from one composition to the next and has incorporated other elements beyond music such as motors, lightbulbs and video.[280]

Onyx Ashanti is a musician and performer who has developed multiple wearable interfaces for musical expression controlling sound synthesis through different sensing technologies integrated with 3D printing technology. Most notable is his BeatJazz instrument, which uses custom hand-held controllers and a custom mouthpiece while connected to a computer to perform live-looping and real-time sound synthesis.[281] and a breathing sensor, inspired by the techniques used in wind instruments. His practice is mostly self-documented and accessible through social media.[282]

The work from these artists highlights the creative potential for music creation found at the intersection of the body and technology where different controllers, movement tracking devices, and wearable technologies are used to perform music. Consumer-grade VR technology offers users many of these technological capabilities where they can explore and prototype movement-based interaction in both virtual and mixed reality environments.

[279] http://www.davidrokeby.com/vns.html accessed January 20, 2025.
[280] https://sonami.net/portfolio/items/ladys-glove/ accessed January 20, 2025.
[281] https://www.youtube.com/watch?v=-0v7mTvJ8M4 accessed January 20, 2025.
[282] https://www.youtube.com/user/onyxashanti accessed January 20,2025.

3.1.2 Implementation

3.1.2.1 Hardware and software platforms

The OpenXR plug-in is built into game development engines such as Unity and Unreal Engine.[283] This plug-in provides a common set of APIs for developing XR applications across multiple VR and AR devices that are commercially available.[284] For this project I worked with Unity[285] and the first release of the Oculus Quest 1 (now discontinued and rebranded as Meta Quest series).[286] Thomas Fredericks' UnityOSC[287] library was implemented to parse tracking and input data within Unity and send it in real-time to Ableton Live[288] via OSC.[289]

3.1.2.2 Movement tracking and controller input

The Quest HMD uses inside-out tracking through its built-in cameras and sensors to track the movement of the hand-held controllers and the user's head, enabling full 6DoF (degrees of freedom) movement. The Unity game engine is used as a 'tracking environment' that enables access to controller input data via scripting. The user's current position and rotation of the HMD and hand-held controllers is measured in reference to the origin in the virtual environment. The position data in Unity is used to compute more complex motion descriptors such as velocity and acceleration with an implementation that adapts some of the methods used in the *modosc* library for marker-based motion capture.[290] The ranges of input data are summarized in Table 1.

---

[283] https://www.unrealengine.com/ accessed January 20, 2025.
[284] https://www.khronos.org/openxr/ accessed January 20, 2025.
[285] https://unity.com/ accessed January 20, 2025.
[286] https://www.meta.com/quest/ accessed January 20, 2025.
[287] https://thomasfredericks.github.io/UnityOSC/ accessed January 20, 2025.
[288] https://www.ableton.com/ accessed January 20, 2025.
[289] Wright, Matthew. "Open Sound Control: an enabling technology for musical networking." *Organised Sound* 10.3 (2005): 193-200.
[290] Dahl, Luke, and Federico Visi. "Modosc: a library of real-time movement descriptors for marker-based motion capture." *Proceedings of the 5th International Conference on Movement and Computing*. 2018.

Name	Type	Value
Primary Button	Discrete	Bool (0 or 1)
Secondary Button	Discrete	Bool (0 or 1)
Thumbstick Button	Discrete	Bool (0 or 1)
Index Finger Trigger (1D axis)	Continuous	0.0 to 1.0
Middle Finger Trigger (1D axis)	Continuous	0.0 to 1.0
Thumbstick (2D axis)	Continuous	-1.0 to 1.0
Rotation (Yaw, Pitch, Roll)	Continuous	(0.0 to 0.5), (0.0 to 1.0), (0.0 to 1.0)
Position (X, Y, Z)	Continuous	-∞ to ∞
Velocity vector (X,Y, Z)	Continuous	-∞ to ∞
Velocity magnitude	Continuous	0.0 to ∞
Acceleration vector (X,Y, Z)	Continuous	-∞ to ∞
Acceleration magnitude	Continuous	0.0 to ∞

Table 1. Data from hand-held controller input.

3.1.2.3 Developing a Max for Live device

The Max for Live (M4L) ‘Connection Kit’ pack[291] has custom devices for sending OSC messages from Ableton Live, but none of these devices can receive OSC messages from external softwares. A custom M4L was developed using the Live API documentation to enable OSC communication and mapping to different parameters in Ableton Live.[292]

291 https://www.ableton.com/en/packs/connection-kit/ accessed January 21, 2025.
292 https://docs.cycling74.com/max8/vignettes/live_object_model accessed January 21, 2025.

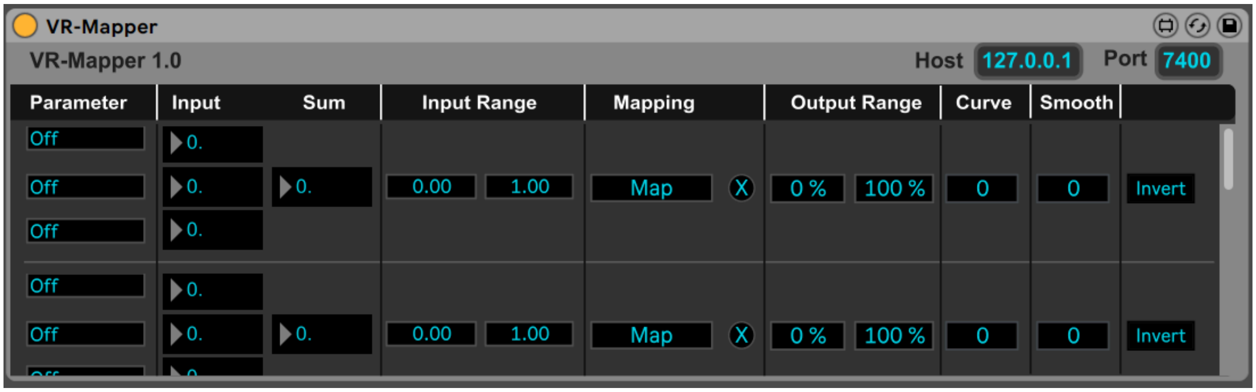


Figure 3. VR-Mapper, Max for Live device.

Hunt and Wanderley, discuss the importance of parameter mapping in electronic instrument design, where they propose moving away from simple one-to-one relationships and work with mappings of higher complexities.[293] They discuss mappings as consisting of 'layers.' In other words, parameters such as 'carrier frequency', 'carrier amplitude', 'modulation frequency' are abstracted into broader sonic descriptors such as 'brightness', 'wobble', 'sharpness', etc. Implementing complex mappings based on many-to-one or one-to-many mappings allows designers to think of concepts like the 'brightness' of a sound as being controlled by multiple parameters together i.e. the cutoff frequency and resonant factor of a filter. In their experiments, the reactions across the test subjects were consistent when exploring complex mappings applied to simple interfaces such as physical sliders. The subjects described these mappings as 'more rewarding' and being 'like an instrument' as they could control several simultaneous parameters without having to 'de-code' them into individual parametric streams.

[293] Hunt, Andy, and Marcelo M. Wanderley. "Mapping performer parameters to synthesis engines." *Organised sound* 7.2 (2002): 97-108.

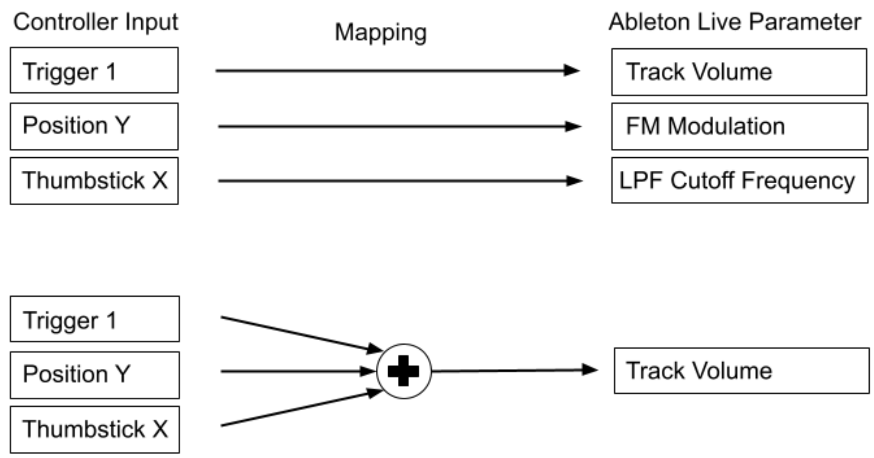


Figure 4. VR-Mapper, one-to-one and many-to-one mappings.

Building on their findings, VR-Mapper enables users to employ one-to-one and many-to-one mappings between user input and the parameters in Ableton. Users can select up to three different data inputs and add them together at a 'node.' The sum is mapped to a Live parameter, as shown in figure 4. Users can adjust the data input range to adjust for the addition of different data inputs. Other features of the M4L device include output smoothing, output inversion, input and output range adjustment, together with linear and exponential options for the output.

3.1.3 *UX-1*, experimental video performance

VR-Mapper was used in the composition of *UX-1*, an experimental video piece that reflects on the phenomenology of online spaces. The name derives from the acronym 'UX' which refers to 'user experience' in the fields of design and human-computer interaction. According to sociologist Viktor Berger, "the use of infocommunication devices conveys the experience of a multiplication, hybridization, fragmentation and multi-layering of spaces—complementing and intensifying the experience of urban, insularized spatiality."[294] We live in a world where the divisions between virtual and real are increasingly less distinguishable for "the realness of objects and living beings is less and less a question of yes or no

[294] Berger, Viktor. "Phenomenology of online spaces: interpreting late modern spatialities." *Human Studies* 43.4 (2020): 603-626.

but more of graduality."[295] The piece was developed during a time when life became increasingly mediated and virtualized due to the COVID-19 pandemic.

The piece combines different audiovisual recordings of myself performing with VR-Mapper. The audiovisual materials were combined in real-time using the open source Open Broadcast Software (OBS) Studio in the same way that a television switch would select from different camera feeds for a live broadcast.[296]

3.1.3.1 Recording movement and sound

For the audiovisual recordings I processed sonic material in real-time in Ableton Live. Pre-recorded vocal material was transformed and processed by mapping movement parameters from the hand-held controllers to various effects and instruments including: granular delay, beat repeater, tremolo and synthesizers. All of the audiovisual recordings were made with a webcam, from the same position I would attend virtual calls in my bedroom. I purposefully made different poses with the idea that these recordings could be overlayed together. Each of the audiovisual recordings consisted of short improvisations with an instance of VR-Mapper.

The background for the video piece is a screen recording of my computer's desktop view, where I preview different files and screenshots for ten minutes (figure 5). These included all sorts of materials: documentation of the VR-Mapper project, digitally processed images made in p5js, quotes from critical theory books, receipts from packages ordered online, calendar appointments, websites, among other random things that existed (and probably still exist) on my computer. The video was saturated in OBS making it easier to apply the chroma key effect for overlays later.

[295] Berger, Viktor. "Phenomenology of online spaces: interpreting late modern spatialities." *Human Studies* 43.4 (2020): 603-626.
[296] https://obsproject.com/ accessed January 10, 2025

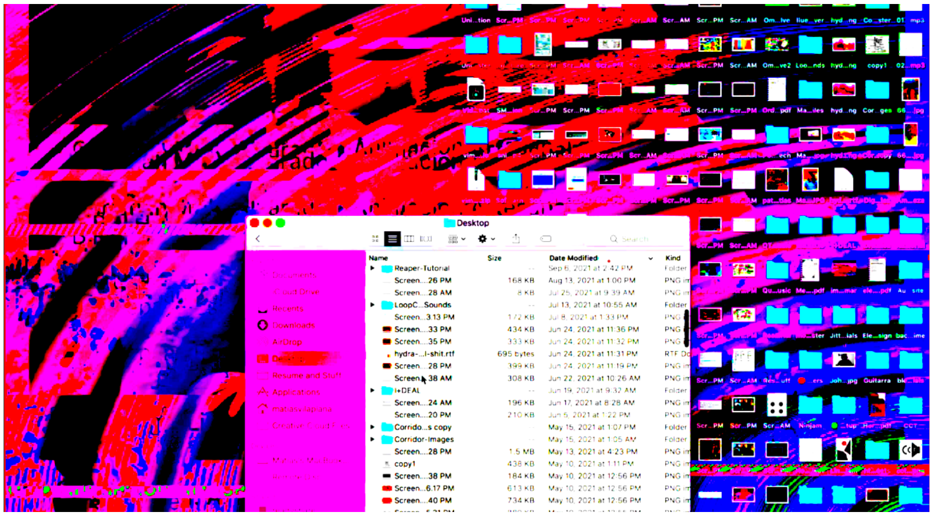


Figure 5. Still frame used in *UX-1* from saturated computer desktop recording.

3.1.3.2 Performing a broadcast software

It is possible to load multiple videos in OBS, putting them together into a playlist from which it is easy to select and playback. The order of the videos in the playlist determines which one has priority in playback. It is analogous to ordering images in a powerpoint, where some can go to the front of the screen and others to the back. The desktop screen recording goes at the bottom, as seen in figure 6, with the audiovisual recordings of myself being put in order above the screen recording. By attaching a chroma key filter to the screen recording video and the performance recordings, it is possible to play one or multiple videos simultaneously while overlaying them together.

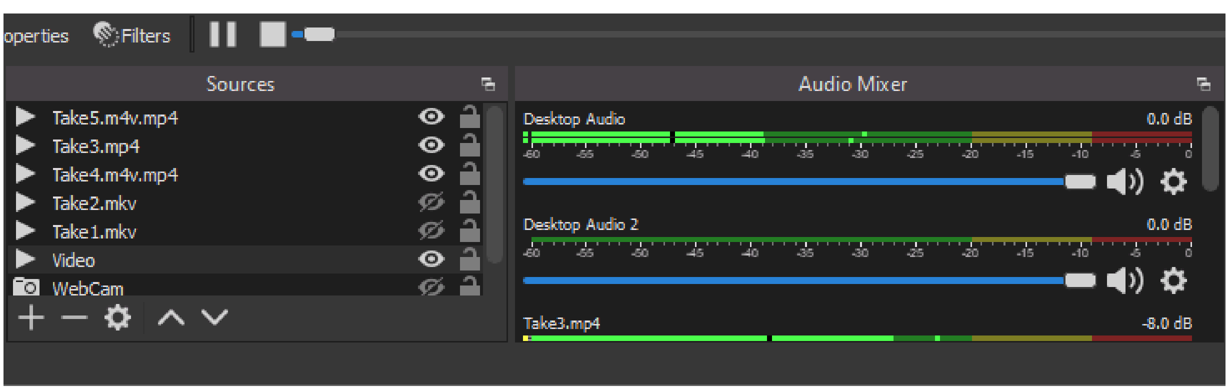


Figure 6. Video recordings organized in OBS.

In this sense, OBS is used as a tool for real-time audiovisual composition, mixing audiovisual performances using VR-Mapper. The piece advances and changes the sonic and visual quality through switching on and off the different audiovisual recordings with no crossfades. The result is a multiplicity of bodies wearing VR headsets and controllers overlayed in granulated and degraded quality (figure 7).[297]

Figure 7. Instances of overlaying audiovisual recordings performing with VR-Mapper.

[297] See the appendix for a recording of the performance.

3.2 *Push & Pull* and *pedalGround*

This section describes the design and implementation of two embodied music interaction systems for VR, *Push & Pull* and *pedalGround*. The two cases presented here consist of a different implementation of the same interaction mode for two different music performances. The first piece, *Push & Pull*, explores the development of a meta-composition for a solo performer immersed in a virtual environment with a HMD and controllers. The second piece, *pedalGround*, explores the interactions between a musician IRL and a VR performer by connecting and mapping interactions between the virtual environment and the MIDI pedalboard of a guitarist.

3.2.1 Design framework, embodied interaction design for virtual environments

In VR users are visually and sonically immersed in a virtual environment. Interaction is possible through a combination of body movements and mappings between the HMDs and hand-held controllers. As explained in Chapter 2, the design of a successful VR experience considers achieving a sense of 'presence' for the user, enabling them to interact with the virtual environment in meaningful ways. Presence in this case is understood as creating the illusion of 'being there.'[298] Achieving this depends on multiple factors, including accurate multimodal feedback to the actions of the user immersed in a virtual environment. However, current consumer-grade technologies are limited in their capacity for full-body tracking and multimodal feedback, especially regarding haptics. This poses limitations to digital embodiment in VR and video games.[299]

Important considerations about embodied interaction design in the context of interactive technologies have been published in the field of HCI. It is relevant to comprehend the role these can play in the design of musical interactions in VR experiences. Dag Svanæs introduces three concepts for embodied interaction design that were inspired in the phenomenological work of Merleau-Ponty; the 'feel

[298] Slater, Mel, and Sylvia Wilbur. "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments." *Presence: Teleoperators & Virtual Environments* 6.6 (1997): 603-616.
[299] Farrow, Robert, and Ioanna Iacovides. "Gaming and the limits of digital embodiment." *Philosophy & Technology* 27 (2014): 221-233.

dimension', 'interaction gestalt' and 'kinaesthetic thinking.' He uses the example of driving a car as an example of the first two:

> "The user experience of having taken the car for a drive is the sum of stimuli in a number of sense modalities: visual, auditory, tactile, and olfactory. In addition, the drive results in the kinaesthetic experience of actually having driven the car. This is the feel dimension of the user experience: how it feels to drive it. It includes how the car reacts on steering the wheel, brake, and gas. The resulting interaction gestalts, such as the experience of operating the manual gear, are not logical composites of action/reaction pair, but atomic percepts in the kinaesthetic sense modality." [300]

Svanæs describes 'kinaesthetic thinking' as our ability to reason about interactive behavior in the 'feel dimension', without having to break down the interaction gestalts into their logical parts or action/reaction couplings. In the design of interactive experiences that require large portions of the body, it becomes relevant to consider the kinaesthetic sense modality, not only as an outcome, but as part of a design process that originates 'from the body.'[301]

The different design frameworks reviewed in Chapter 2 suggest the development of 'magical interactions' for the design of musical instruments in VR, where the design of an interaction will "qualify as magical if it is not limited by real-world constraints, such as the ones imposed by the laws of physics, human anatomy, or the current state of technological development."[302]

An emphasis on developing 'magical' or 'impossible' embodied interactions, particularly through the idea of 'kinaesthetic thinking', has been the cornerstone for the design of this musical interactive system that uses body movements to control virtual objects through simulated gravitational forces in the form of telekinesis.

---

[300] Svanæs, Dag. "Interaction design for and with the lived body: Some implications of Merleau-Ponty's phenomenology." *ACM transactions on computer-human interaction (TOCHI)* 20.1 (2013): 1-30.
[301] Svanæs, (2013).
[302] Serafin, Stefania, et al. "Virtual reality musical instruments: State of the art, design principles, and future directions." *Computer Music Journal* 40.3 (2016): 22-40.

3.2.2 Implementation

3.2.2.1 Hardware and software platforms

Both of the interactive systems use an Oculus Quest 1 tethered to a PC running the Unity game engine. The sound interactions in both of the virtual environments are determined by collisions between 3D spheres and other virtual objects. OSC is used to communicate the collision data to Max/MSP and Ableton Live. In *Push & Pull*, the collision data between the spheres and other virtual objects is sent via OSC to a custom M4L (Max for Live) device in Ableton Live to trigger the onset of a MIDI instrument (figure 8a). In *pedalGround*, the collision data trigger MIDI messages which are sent to a MIDI pedal board via OSC in order to control the on/off state of different analog effects (figure 8b).

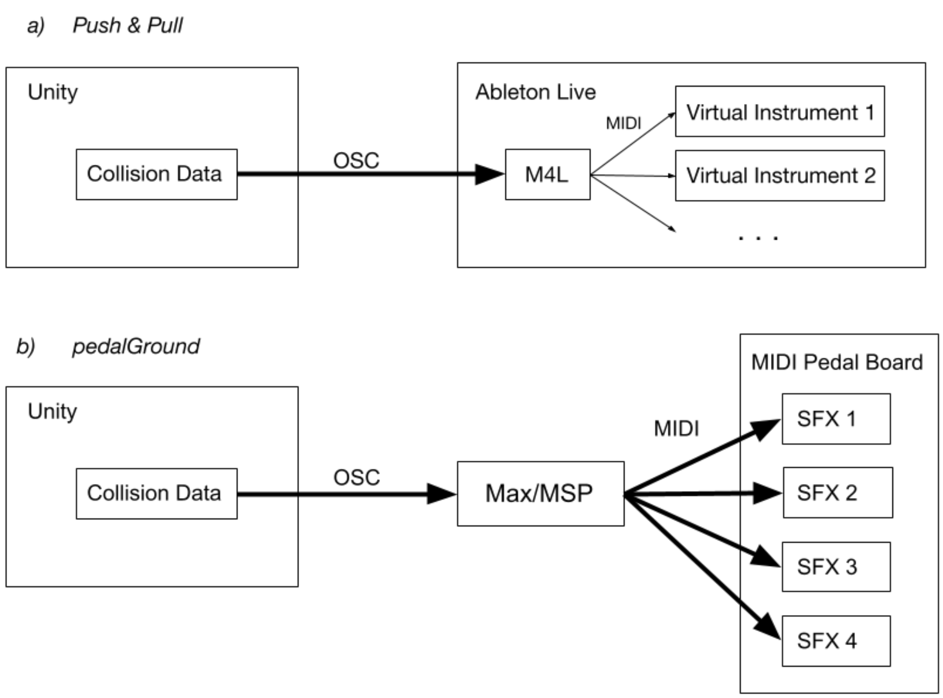


Figure 8. Data structure for *Push & Pull* (a) and *pedalGround* (b).

3.2.2.2 System mechanics and controller mapping

One of the advantages of game development engines such as Unity, is the capacity to customize the physics of a virtual environment. Virtual objects can be assigned components such as ‘colliders’ and ‘rigidbodies’ to simulate physical interactions such as collisions and forces between different objects. I consulted Daniel Shiffman’s book, Nature of Code,[303] to learn how to model different types of forces found in the natural world such as wind, friction and gravitational force, together with more general outlines on how to model any forces one could envision in the p5js environment. These models were later ported into Unity’s scripting language, C# code, and were further customized for three dimensions.

The mechanics of the interactive systems are based on Newton’s law of universal gravitation,[304] where the force of attraction between two masses is directly proportional to the product of their masses and inversely proportional to the square of the distance between their centers, as shown in the following equation:

$$F = G \frac{m1 * m2}{r^2}$$, r = distance between masses; m = mass, G = gravitational constant

The interactive system developed enables performers to use Newton’s law as a form of ‘telekinesis’ through attraction and repulsion forces to control an interactive musical system in VR. In the virtual environment the hand-held controllers become points of gravitational force towards which the 3D spheres are attracted. A total of eight virtual spheres of different masses are attracted towards each of the virtual hands. The vector of the gravitational force is always pointing from the center of each sphere towards each of the hands, as shown in figure 6. The magnitude and direction of the force vector is updated and applied on every frame to each of the spheres. Through timed gestures, the user can change the distance between the hands and the spheres, setting the latter into motions of elliptical nature around the virtual hands. The length of the ellipses or how far and close they move from the point of gravitational force will depend on the timing and amplitude of the hand gestures, making it possible to ‘swing’ the spheres

[303] https://natureofcode.com/ accessed January 27, 2025
[304] https://imagine.gsfc.nasa.gov/features/yba/CygX1_mass/gravity/more.html accessed January 27, 2025

around the user at long distances, or to 'juggle' the spheres with movements of shorter distances. The collisions between these spheres and other virtual objects are sonified and/or used to trigger other sound processes as exemplified in the two examples described later in this section.

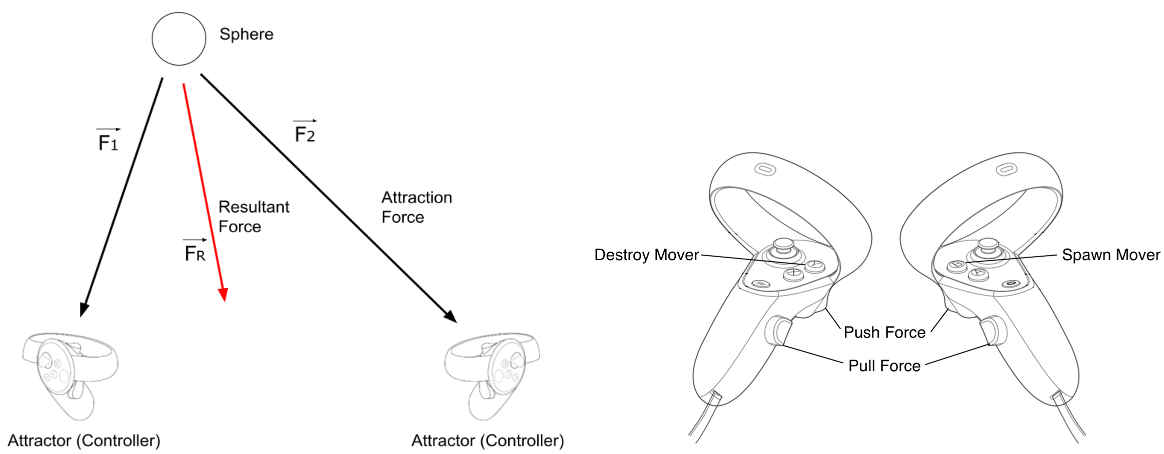


Figure 9. (a) System physics mechanics. (b) Controller input.

Using the button input on the handheld controllers the user can control the amount of spheres in the virtual environment and also apply external forces on the spheres, as shown in figure 9a. The primary buttons can be used to spawn or destroy spheres, with a limit between zero and eight spheres to be present in the virtual scene (figure 9b). The triggers in the controllers are used to apply 'push' and 'pull' forces on the spheres, giving the user 'telekinetic' agency over the spheres while overlapping with the permanent attraction of the spheres towards the hands. The 'push' force is applied in the forward direction in which the controllers are pointing, while the 'pull' force is always applied in the direction from the sphere's center towards the hands.

3.2.3 *Push & Pull*

In this virtual environment the user is floating above the ground in a barren landscape with steep mountains on the horizon as shown in figure 10a. The user is surrounded by three concentric layers of virtual objects (which we will call targets) as shown in figure 10b. There are four types of targets in the scene – differentiated by shape, color and texture as shown in figure 11a. Each of these targets is mapped to a virtual MIDI instrument. When a sphere collides with any of the targets, the target will light up and trigger a note onset to its corresponding instrument in Ableton Live. The spheres also have an associated instrument that is only triggered when they collide with each other.

Figure 10. (a) Landscape in *Push & Pull*. (b) Top view of spatial arrangement of virtual objects.

Figure 11. (a) Target virtual objects in *Push & Pull*. (b) Different sized spheres controlled by the user.

The virtual instruments chosen include a synthesizer, violin, cello and percussion for the targets, and a piano for the collision between spheres as shown in table 2. The pitch is determined by the mass of the sphere that collides with the target, while the loudness and duration of the notes triggered are determined by the speed of the sphere at the moment of collision. The pitch association between each sphere and targets is summarized in table 3.

Virtual Object	Virtual Instrument
Spheres (1-8)	Piano
Target 1	Synthesizer
Target 2	Cello (legato)
Target 3	Violin (pizzicato)
Target 4	Percussion

Table 2. Virtual object and virtual instrument mapping.

Using the controllers, the user can combine their hand movements together with applying push/pull forces to direct the trajectory of the spheres towards the targets surrounding them. Various constraints were designed into the environment to increase the difficulty of hitting the targets. The mass differences between the spheres makes the resultant forces different for each sphere. This produces differences in the motions and trajectories between each sphere, making it difficult for the user to predict and direct the movement of all the spheres simultaneously. The targets themselves present a constraint to the sphere's motion as well by either completely stopping their movement or redirecting it. Using a script with simple periodic oscillation, some targets are set to oscillate their position along an axis at a predetermined amplitude and speed, adding a timing difficulty to the task of hitting the targets.

Sphere #	Piano	Synthesizer	Cello (legato)	Violin (pizz)	Percussion
1	D3	D2	D2	D2	D2
2	D4	D4	A2	A2	A2
3	A4	A4	D3	D3	D3
4	F5	F5	E3	F3	G3
5	A5	A5	F3	A3	D4
6	D6	D6	A3	D4	A4
7	A6	A6	D4	F4	F5
8	D7	D7	F4	D5	A5

Table 3. Pitch mapping between spheres and virtual instruments.

3.2.3.1 Composing space

By design the system can be thought of as a meta-composition, as it enables musical interactions without a determined time constraint while simultaneously distributing the agency between the performer and the virtual physics that govern the system.[305] For a particular instance of this meta-composition, the piece is structured by exploring the different ranges the spheres can be pushed to, continually expanding the egocentric space of action in the performance through the interactive mechanics of the system.[306] Different pieces could be composed by changing the number and type of targets, with different spatial arrangements. This would in turn change the gestures from the user in order to exploit the full possibilities of the system.

[305] Bahn, Curtis Robert. *Composition, improvisation and meta-composition*. Princeton University, 1997.
[306] See the appendix for a recording of the performance.

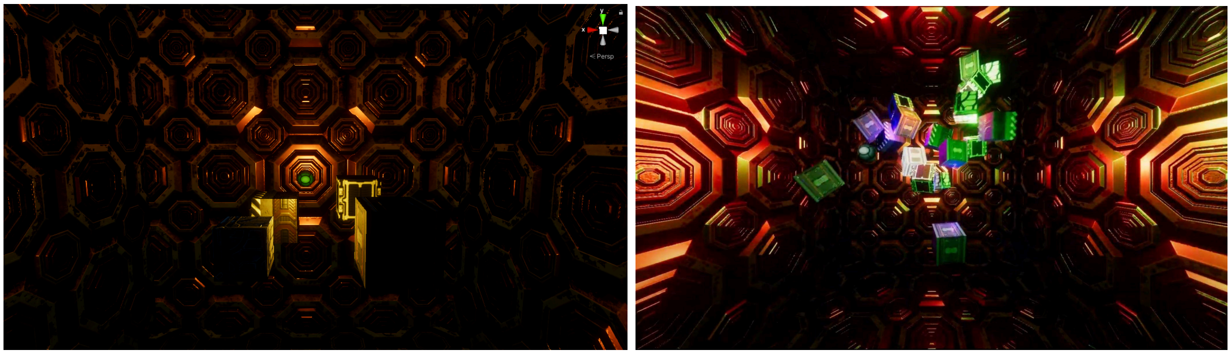


Figure 12. View of the virtual environment in *pedalGround.*

3.2.4 *pedalGround*

The mechanics for this virtual environment are built upon those developed for *Push & Pull*, using gravitational forces and telekinetic interactions. The virtual environment (figure 12) is an enclosed space, in the form of a hollow cube with the user floating next to a lateral wall. Initially, four different cubic shaped targets float around the scene, they are differentiated by color and texture as shown in figure 13a. The VR performer can determine the amount of spheres in the scene and can control the movement of the spheres through the same telekinetic interactions as the ones described for *Push & Pull*. The inside of the hollow cube is lit up by eight audio reactive point lights. These lights move around the center of the cube in different periodic motions. Besides emitting light, these points also function as attractors that pull the targets towards them, producing the targets to move in random motions. The enclosed space constraints both the movers and targets from moving outside of the volume, often hitting and bouncing off the walls.

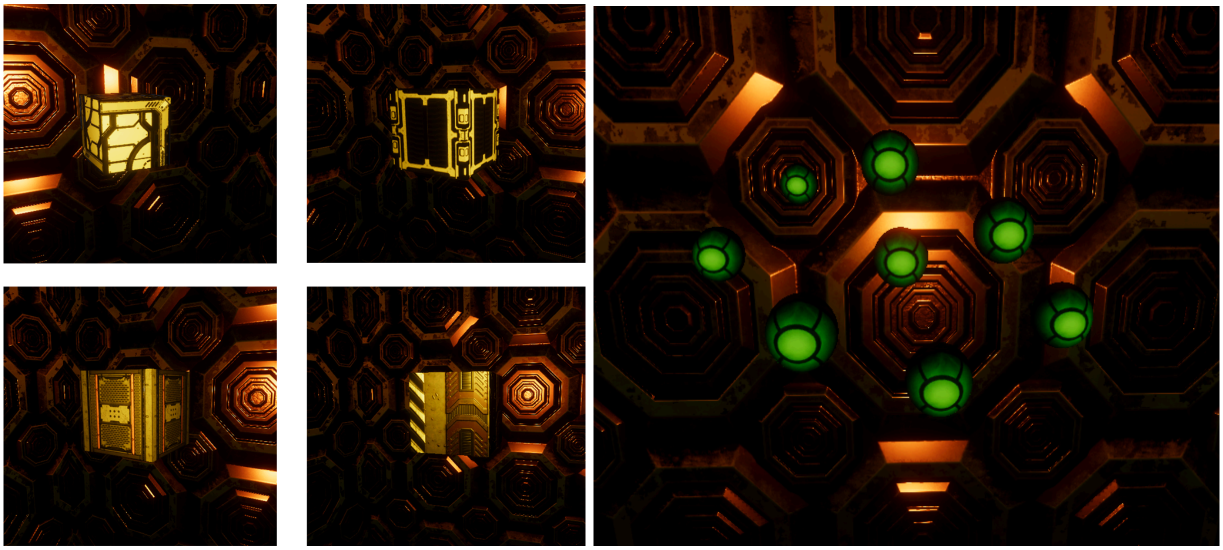


Figure 13. (a) Target virtual objects in *pedalGround*. (b) Different sized spheres controlled by the user.

3.2.4.1 Mixed-reality interactions

In this iteration, the VR system is designed to interact with the MIDI pedalboard of a guitarist in the context of a live performance. The sound is produced by the guitarist, while the actions that take place in the virtual environment change the switches in the MIDI pedalboard of the guitarist. In order to establish a bi-directional interaction the sounds produced by the guitarist are mapped to control the intensity and colors of the light emitters in the virtual environment. Both the guitarist and virtual performer hear the same sonic output, while the first person view of the user in the virtual environment is projected onto a screen for both the guitarist and audience to see. This configuration is summarized in figure 14.

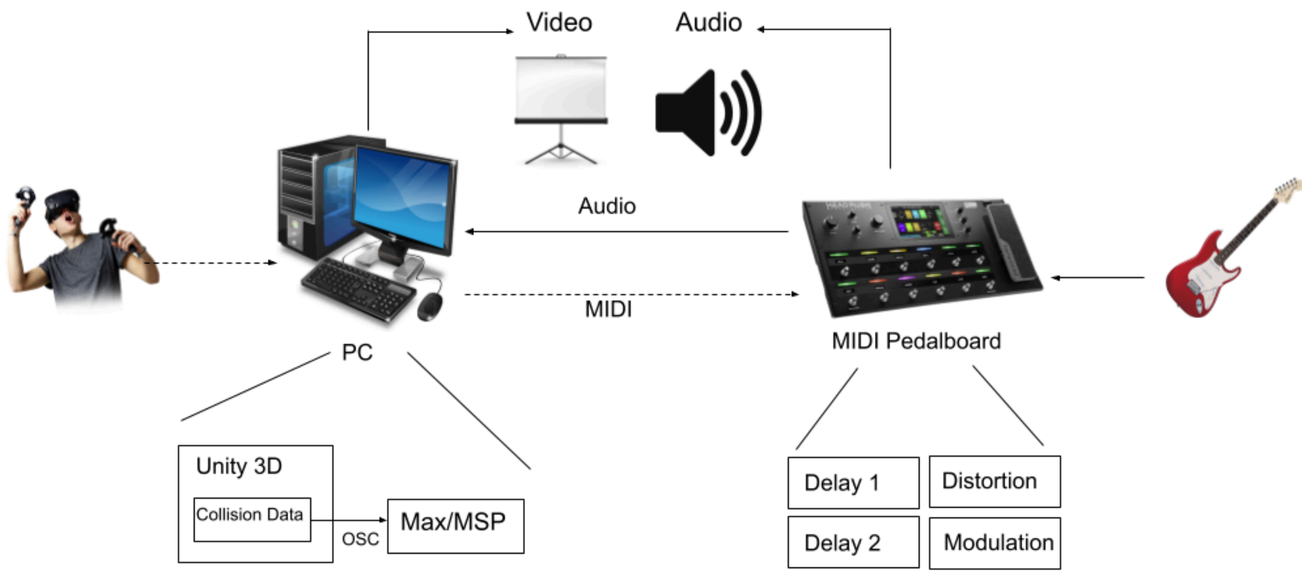


Figure 14. System configuration for mixed-reality interactions in *pedalGround*.

In this configuration, the collision between spheres and targets sends MIDI messages to the guitar pedalboard, switching the on/off state of different effect pedals. We decided to use effects that would produce a notorious change in the sound quality regardless of their combination. Each effect was mapped to a specific target in the virtual environment, which would be toggled 'on' or 'off' whenever hit by any of the spheres, as shown in table 4. All other collisions – between the spheres themselves and or walls – do not have any consequences on the sound interaction.

Virtual Object	SFX Pedal
Target 1	Distortion
Target 2	Modulation
Target 3	Analog Delay
Target 4	Digital Delay

Table 4. Target to audio effect mapping.

3.2.4.2 Gamified music improvisation

After a few rehearsals where we improvised together with the guitarist, we brainstormed ideas on how we could structure the performance with our musical system. We identified the difficulty involved in hitting the moving targets with the spheres using the telekinetic interaction. We experimented with incrementing the number of targets in the scene, increasing the probability of one of the eight spheres colliding with the targets. The result was an increased rate of change of the on/off states of the effect pedals. We decided to structure the musical performance around the idea of 'rate of change'. The pseudo-random motion and the populatedness of the scene can be thought as the constraints for the user's capacity to hit the targets. A gamified mechanic was implemented in order to dynamically increase and decrease the number of targets and how they moved in the virtual scene, affecting the overall rate of change in the performance. We achieved this by implementing two game 'states' in the virtual environment. The VR performer can select between the two states using the controller joystick.

Game State 1. When a sphere collides with a target, a target of the same type is spawned in the scene. This is true for all four types of targets, with a capacity of eight targets per type, with a total capacity of thirty-two in the virtual scene.

Game State 2. When a sphere collides with a target, that target is instantly destroyed. The targets on the scene can be reduced to only four, with one of each type.

Movement State 1. In this state the movement of the targets has a gliding and expansive quality, accelerating only when the targets are close to the light emitters. The attraction force of the light emitters on the targets is inversely proportional to the distance between them, as represented in the following equation:

$$F = G \frac{m1 * m2}{r^2}$$

Movement State 2. In this game state the movement of the targets has a thrusting quality, clustering towards the light emitters, with the cluster often moving around as a single unit. The attraction force increases proportionally to the distance between the targets and the light emitters, as represented in the following equation:

$$F = G r^2 (m1 * m2)$$

The musical performance took place at The Bridge Arts Initiative in Charlottesville, VA. The guitarist laid down the guitar and utilized an e-bow in order to perform sustained sounds of drone-like quality, while the VR performer selected between the different game and movement states. Both the tasks of aiming towards targets and avoiding them are at play in the combination of motion and game states. Simultaneously, the guitarist chooses to either react to the effect changes or to maintain their sonic trajectory. The musical improvisation is in dialogue with the visuals from the virtual environment through audio reactive lighting in the scene. The structure of the improvisation oscillates between high and low states of visual and sonic density.[307]

[307] See the appendix for a recording of the performance.

3.3 *Unforeseen Collisions:* music collaboration in shared virtual environments

Musical collaboration in virtual environments presents multiple challenges and opportunities. People have been interacting together in virtual environments for many years now, most commonly in the form of multiplayer online gaming. In recent years, there has been an increase in VR applications that enable users to engage in musical activities, including tools for music creation and creative arrangement, rhythm games, and VR performances.[308] However, most of these applications are for single users and do not support multi-user interactions. It is surprising since musical creativity is oftentimes a collective experience; we make music in presence of others and perform with others as well. On the other hand, networked music is a field that already deals with numerous challenges in system design and performance aesthetics.[309] [310] Introducing VR into networked music presents additional challenges in terms of the synchronization and communication between the performers in different modalities, requiring a careful evaluation and optimization of the system state to determine which aspects of it should be shared over the network at any given moment.

In the following sections I will describe the design and implementation of *Unforeseen Collisions*, a networked VR system for musical collaboration, and discuss the creation of a musical composition for two networked performers and the challenges involved in presenting this work to an audience. The system is a large-scale virtual environment that facilitates the collaborative design of musical causalities for multiple users over a network. Leveraging the physics engine built into Unity, the system allows performers to create intricate audiovisual polyrhythms by placing resonant blocks in the pathway of a constant stream of marbles.

---

[308] Loveridge, Ben. “An overview of immersive virtual reality music experiences in online platforms.” *Journal of Network Music and Arts* 5.1 (2023): 5.

[309] Rottondi, Cristina, et al. “An overview on networked music performance technologies.” *IEEE Access* 4 (2016): 8823-8843.

[310] Lemmon, Eric C. “Telematic Music vs. Networked Music: Distinguishing Between Cybernetic Aspirations and Technological Music-Making.” *Journal of Network Music and Arts* 1.1 (2019): 2.

3.3.1 Design Framework

The design of *Unforeseen Collisions* was informed by three guidelines:

> 1) The interface should only be possible to implement in VR, exploiting interactive, virtual-physical, acoustical, and visual affordances of VR.
>
> 2) The interface can be governed by existing musical traditions, mental models and learned sensorimotor behavior or deviate from these entirely in the form of a novel interface.
>
> 3) The scale of the interface can range from that of a hand-held instrument to that of an arbitrarily sized environment conceived as a musical system.

These guidelines stem from previous work on developing Virtual Interfaces for Musical Expression (VIME), where we evaluated how various design considerations can affect the user experience in terms of control, physical effort, and immersion, among other factors.[311] These guidelines portray similar principles to the design frameworks described in Chapter 2. For example, the design of 'magical interactions' should not be "limited by real-world constraints, such as those imposed by the laws of physics, human anatomy, or the current state of technological development"[312] and to "make things that would be impossible in the physical world"[313] resonate with guideline 1). Furthermore, the frameworks described earlier highlight the significance of virtual embodiment and music as a social experience.[314] [315] Accordingly, *Unforeseen Collisions* explores the role of co-presence and collaboration in a musical VR experience, while at the same time serving as a platform with which to build a performance practice in VR.

[311] Çamcı, Anıl, Matias Vilaplana, and Ruth Wang. "Exploring the affordances of VR for musical interaction design with VIMEs." *Proceedings of the International Conference on New Interfaces for Musical Expression*. 2020.
[312] Serafin, Stefania, et al. "Virtual reality musical instruments: State of the art, design principles, and future directions." *Computer Music Journal* 40.3 (2016): 22-40.
[313] Atherton, Jack, and Ge Wang. "Doing vs. Being: A philosophy of design for artful VR." *Journal of New Music Research* 49.1 (2020): 35-59.
[314] Serafin et al. (2016).
[315] Atherton and Wang. (2020).

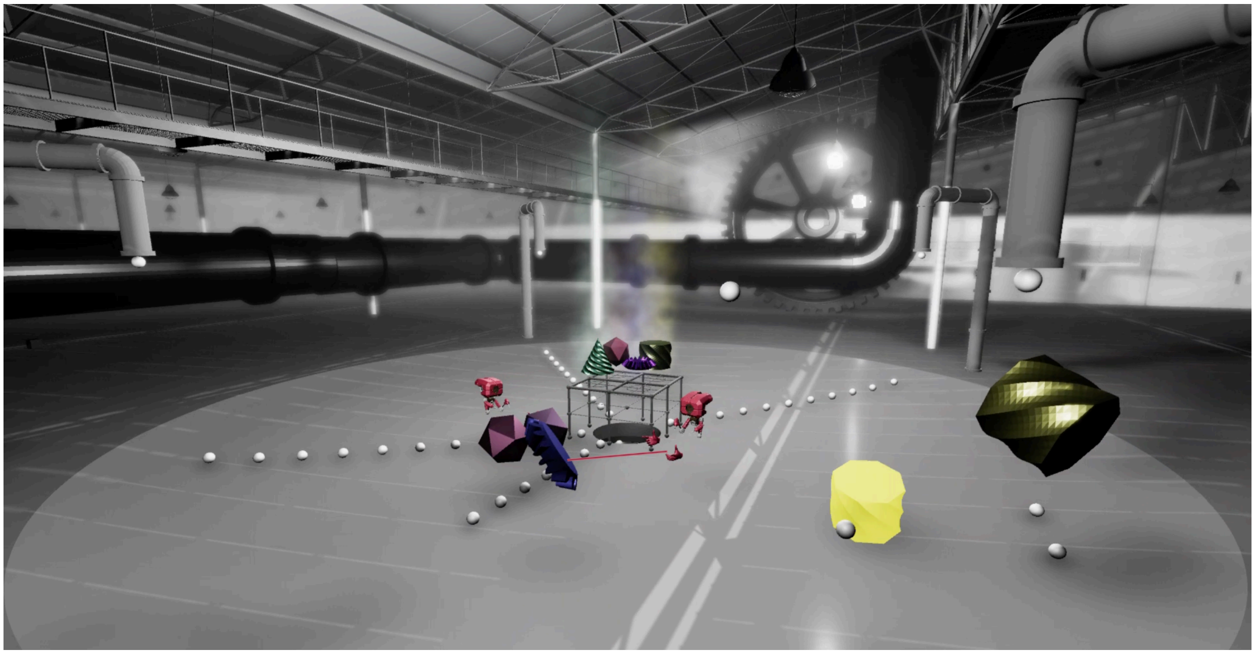


Figure 15. Virtual camera pointing towards the two performers in the *Unforeseen Collisions* play area.

3.3.2 Virtual environment and system mechanics

The performers are situated in a virtual environment that resembles a factory with a play area at its center, as seen in figure 15. The virtual environment is room-scale, performers can navigate this play area by either moving in physical space or virtually by using the joystick on one of the hand-held controllers. At the four corners of the play area are four pipes that extend from the floor to the ceiling and curve back towards the ground. While the system is running, an endless stream of marbles are spawned at an adjustable rate from each pipe. Once a marble falls on the floor, it rolls towards a drain placed at the center before it falls off the scene.

At the center of the play area, four blocks of different shapes are placed on a pedestal, as shown in figure 16. When a performer grabs one of these blocks, a new instance of the same block is created in its place, allowing performers to spawn an arbitrary number of blocks from the pedestal. The blocks can be placed anywhere in the room and will float in midair since they are not affected by gravity. When a block is struck by a marble, it gives off a sound and lights up. Each block shape is assigned a custom modal synthesizer with a dedicated bell or glass-like sound.

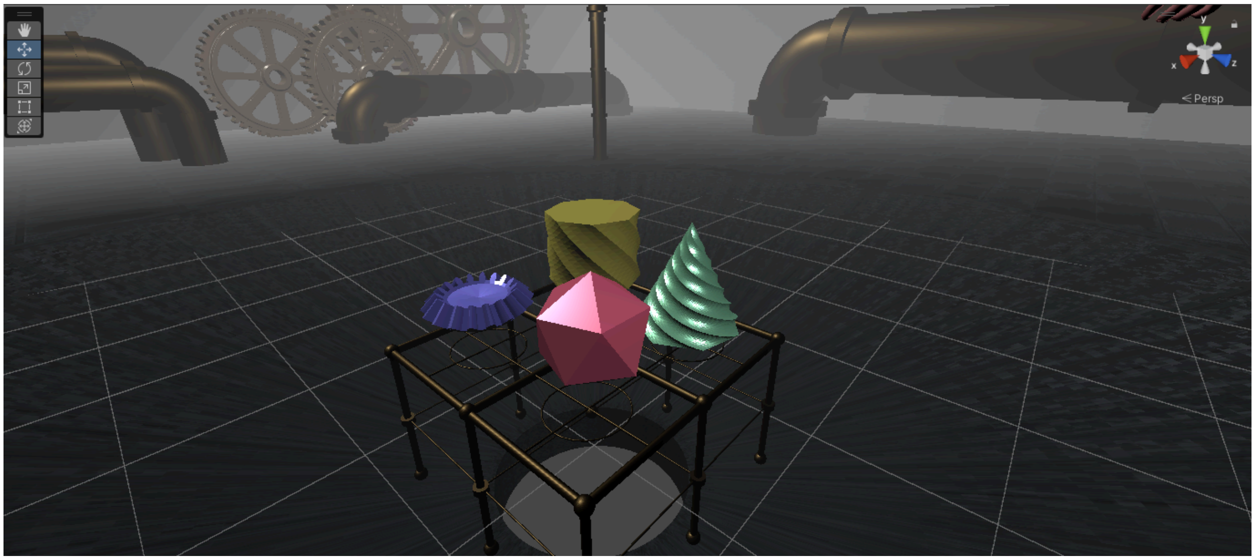


Figure 16. Resonant blocks on a pedestal in *Unforeseen Collisions*.

A marble that collides with a block will bounce off in a direction determined by the shape and placement of the block. The velocity of the collision between a marble and a block determines the loudness of the sound emitted. By placing various blocks in the pathway of a stream of marbles, the performers construct a musical causality system, where marbles bounce from one block to another until they fall off the play area through the drain. Through this physics-based mechanic, the performers can create audiovisual polyrhythms. While physics collisions are enabled between the marbles and blocks, they are disabled among the blocks themselves. As a result, multiple blocks can be overlapped in virtual space, allowing the creation of compound blocks that can play chords when struck by a marble.

Blocks floating in mid-air can be repositioned by grabbing them. By holding a block or laser-pointing at a distant one, the performers can also destroy, mute, or pitch-shift the block. The pitch of a block can be changed in a pentatonic scale using the primary buttons on the hand-held controllers. The pitch of a block can also be bent continuously at a microtonal level using the joysticks while laser-pointing at it. While the pentatonic quantization allows the performers to remain in tune without extensive effort, the microtonal bending allows performers to achieve more complex scale relationships.

In a similar fashion, the performers can laser-point at a pipe to control the spawn rate of the marbles emitted from it, effectively setting the tempo for the musical causality chains associated with that specific pipe. Using the primary buttons, the performers can increase and decrease the spawn rate of a pipe by steps of 15 beats per minute (BPM) within a range from 15 to 120 BPM.

3.3.3 Hardware and Software Implementation

Unforeseen Collisions was implemented in Unity and can be loaded on the Meta Quest platform. At the time of the performances the users were tethered to computers running Unity via Meta Quest Link cables. In the latest version the system can run in on a standalone Quest 2 and Quest 3 headset. The sounds that are emitted from the blocks are spatialized using Resonance Audio,[316] providing binaural output from the system. The computer network between the performers was implemented using the free version of the Photon Unity Networking (PUN) API.[317] This version provides server access to rooms with a limited number of users within the same geographic region.

3.3.4 Network Design

Some of the elements in the system, such as interactable objects and player avatars, need to be synchronized over the network, whereas other elements, such as static meshes and particle systems remain unsynchronized as local copies. The synchronization of dynamic elements is achieved by setting generic *GameObjects* in Unity as networked objects utilizing the *Photonview* component from the PUN API. This allocates a dedicated ID to an object, allowing its state parameters, such as its transform properties (i.e., position and rotation), to be synchronized over the network.

In *Unforeseen Collisions*, the blocks are set as network objects, whose ownership defaults to the performer who instantiates them. Their states are broadcast over the network only when a performer

[316] https://resonance-audio.github.io/resonance-audio/ accessed January 28, 2025
[317] https://www.photonengine.com/pun accessed January 28, 2025

interacts with them. When a performer touches or points the laser at a block, its ownership is transferred, allowing the performer to change its properties (e.g., pitch, mute state) and broadcast them over the network. This prevents the two instances of the system from going out of sync if two performers attempt to apply conflicting changes to a block at the same time. When a performer grabs a block, the synchronization messages are sent in each frame to accurately display the current transform of the block to the other performer. A similar approach is followed with the performer avatars, where their transform properties are broadcast in each frame, allowing performers to gain a clear sense of what the other performer is doing at any given time.

3.3.5 Live music performance with *Unforeseen Collisions*

In this section, I discuss two performances with *Unforeseen Collisions*: an online performance at an international conference and a hybrid performance at a music festival.

3.3.5.1 Online Performance

This performance took place at a conference during the COVID-19 pandemic, the event was held online much like many other conferences at the time.[318] The performers joined the virtual environment from their respective homes, using a VR headset tethered to a computer running Unity. We used a third computer to broadcast the performance over Zoom for the conference attendees. On this computer we set up a virtual camera on a third Unity network to present a third person perspective of the virtual environment. A listener node was also placed on this camera for sounds to be spatialized in reference to it, enabling an accurate depiction of this perspective for the audience. The camera was pointed at the play area and rotated around the center, as shown in figure 13.

[318] See appendix to find a link to the performance.

Figure 17. Hybrid performance. In-person performer (left). Remote performer (right).

3.3.5.2 Hybrid Performance

We did a second performance with the system using a hybrid configuration, with one of the performers on stage while the second performer joined from a remote location (figure 17). It was presented at the Digitalis ‘22 Electronic Music Festival at the University of Virginia in May 2022.

In the concert hall, the in-person performer used two computers: one to drive the HMD and interact in the virtual environment, while the second computer was used to project the virtual environment on a screen for the audience to view. The third-person perspective from this second computer was set to a static position in the virtual scene with the sound spatialized in reference to the position of the virtual camera. The binaural render from the camera view was connected to the stereo playback system in the concert hall. A diagram that describes the system setup for this performance is shown in figure 16.

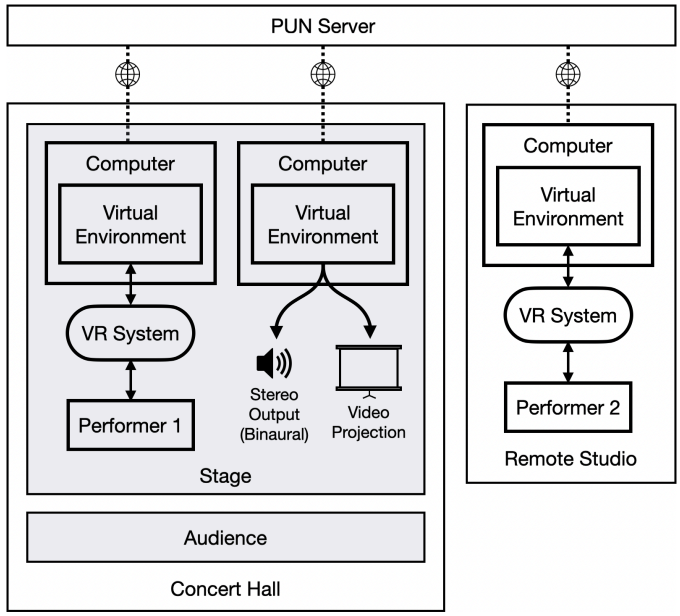


Figure 18. Technical setup for hybrid performance.

3.3.6 Musical improvisation and composition in *Unforeseen Collisions*

While iterating on the system's design, we carried out numerous rehearsals with a collaborator to test the network features of the system. This included free improvisation sessions that led us to discover musical affordances we had not anticipated while designing the system. For instance, the discovery of overlapping blocks to generate chords or how to concatenate blocks in a way that marbles could slide down to create arpeggios. While experimenting with these ideas, we also recognized slight inconsistencies in the game engine's collision calculations. When blocks were stacked on top of each other, marbles would bounce off in different directions, adding a degree of indeterminacy that enabled us to develop richer rhythmic and melodic articulations. This resulted in a less deterministic system with a more natural behavior, giving the system a degree of randomness that we would expect in the physical world.

When we attempted more structured compositional approaches, these revolved around the idea of construction and deconstruction. This involved developing rhythmic patterns that grew in density, to later trace our way back and deconstruct these patterns by repositioning, muting, or destroying the blocks. We developed different spatial choreographies in the form of written instructions that would lead to new patterns, changes in the density of existing patterns, and moments of silence. The following lines are an example of the instructions for a spatial choreography used in a live performance:

Spatial Choreography
Begin in opposing corners
Build a structure with 3 blocks
Move to the corner to your right
Build a structure with 3 blocks
Move to the corner to your right
Reduce to one block by muting 2 of the existing blocks
Move to the corner to your right
Reduce to one block by muting 2 of the existing figures
Play with that block (and the pipe as you see fit)
Build a bridge to the structure to your right with 3 blocks maximum
Move to the corner to your right
Build a bridge to the structure to your right with 3 blocks maximum
Mute all blocks down to 1 (we shouldn't take turns for this, do in a slow manner, play by ear)
Play with that one block for a moment
Start turning on blocks slowly while playing with them
When the music gets complex enough, start turning every block on rapidly
Move back to your beginning pipe
Move the block at the origin to change the trajectory of marbles
Move to the corner to you right and move the block at the origin to change the trajectory of marbles
Play around with rhythm (play by ear)
Destroy the block at the origin
Move to your right and destroy the block at the origin to end the piece.

Although this performance sequence led to a playful interchange between the performers, it also meant that the two were always spatially separated in the virtual environment. This prompted us to explore ways to further leverage the virtual space as a shared environment. For our second performance we identified strategies that would clearly display the virtually co-located nature of the performance to the audience. These involved the two performers constructing structures together or having one performer position blocks while the other one is tuning them or adjusting the tempo associated with that structure.

3.4 Virtual environments as graphic scores for music improvisation

In Chapter 2 I described creative works in XR that exemplify the new opportunities these technologies present for the development of novel musical instruments and interfaces. However, these frameworks and examples shed little light on how these technologies can interact with already existing musical practices, such as creating scores or performance systems for musicians playing traditional instruments. But how can XR challenge and expand the musical creativity of instrumentalists, and improvisers without needing them to learn a new instrument? In Chapter 1 I described the example of George Lewis' *Voyager*, where the program he developed is the virtualization of an improviser who can interface with other musicians, challenging them in how they think about improvisation. Following that example, XR can be thought of not only as a medium to perform music, but also as a medium to organize musical information for others to perform. *Resilience* by Jack Atherton (Chapter 2) is a good example of this, where there is a VR conductor using both their gestures and the virtual environment to direct the musical performance of a laptop ensemble.

3D virtual environments present opportunities for the design of 3D graphic scores, transcending the two-dimensional page upon which music is traditionally notated. Researchers have investigated how users can create 3D graphic scores for screen-based non-immersive virtual environments and study the creative practices that have emerged from it.[319] Composer Kim-Boyle has explored the representation of musical form using generative 3D scores.[320] In his works *point studies no. 2*, *64x4x4*, and *5x3x3* he uses 3D structures built of colored nodes connected by thin white lines to present performers with different pathways representing pitch and duration. The camera view constantly shifts to present performers different pathways they could interpret. Kim-Boyle acknowledges the limitations of presenting a 3D environment on a 2D screen and has explored porting some of his work onto a mixed reality headset,

[319] McCall, Lauren, and Jason Freeman. "A 3D Graphic Score Space and the Creative Techniques and Performance Practices that Emerge From It." *Proceedings of the 16th International Audio Mostly Conference*. 2021.
[320] Kim-Boyle, David. "The 3-d score." *TENOR 2017: International Conference on Technologies for Music Notation and Representation:[24-26 May 2017, University of A Coruña, Spain]*. Facultade de Filoloxía, 2017.

although this presents other challenges for composers in arranging physical space for performers to interact with scores presented in mixed reality.[321]

3.4.2 *Senderos*: A dynamic 3D graphic score

The dynamic 3D graphic score for this composition was created using satellite images that were later rendered as 3D virtual landscapes. The 3D landscapes were imported into a virtual environment and spatially arranged. A virtual camera was automated to hover over the different landscapes providing a 'bird's eye view' of the landscape. As the camera navigates over each of the landscapes, performers are given a separate score outlining how different musical parameters are associated with each landscape enabling them to musically interpret what they see on the screen.

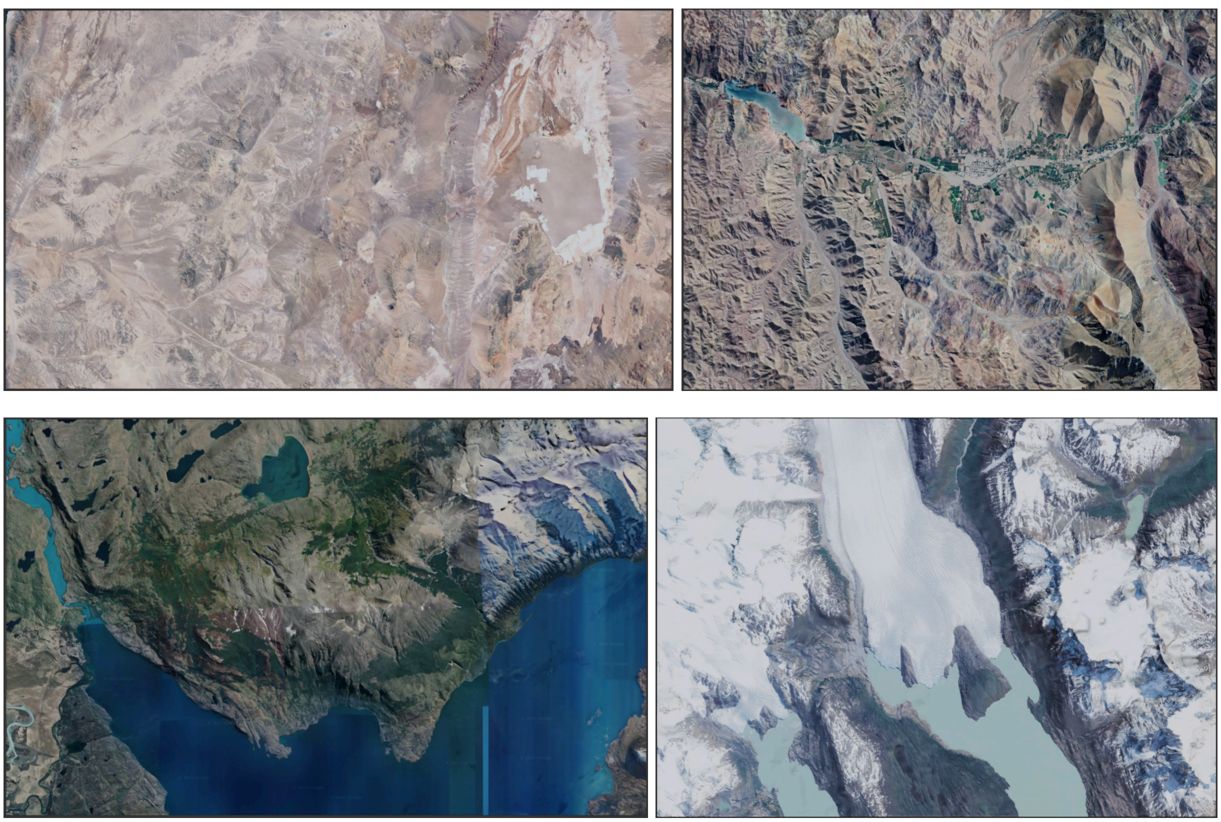


Figure 19. Satellite images used for *Senderos*.

[321] Kim-Boyle, David. "3D notations and the immersive score." *Leonardo Music Journal* 29 (2019): 39-41.

3.4.2.1 Implementation

Using Blender, an open source 3D modeling application,[322] satellite images were imported and rendered in three dimensions. It is possible to access Geographic Information System (GIS) data from multiple sources with the BlenderGIS plug-in.[323] The elevation data from the satellite images is used to render the landscapes in three dimensions. It is possible to include meshes to represent buildings of a particular location by pairing the GIS data with data from OpenStreetMap.[324]

Figure 19 shows the satellite images of the locations selected for the graphic score. All four of them were imported into the Unity game engine to be scaled and spatially arranged as shown in figure 20. A virtual camera was automated to follow a predetermined trajectory in the virtual environment using the free 'Camera Path Creator' add-on in Unity.[325] Other parameters for automation include speed and rotation of the virtual camera. The path determined for the camera goes through all four landscapes. The video component of the 3D graphic score is a recording from the view of the camera hovering over the landscapes.

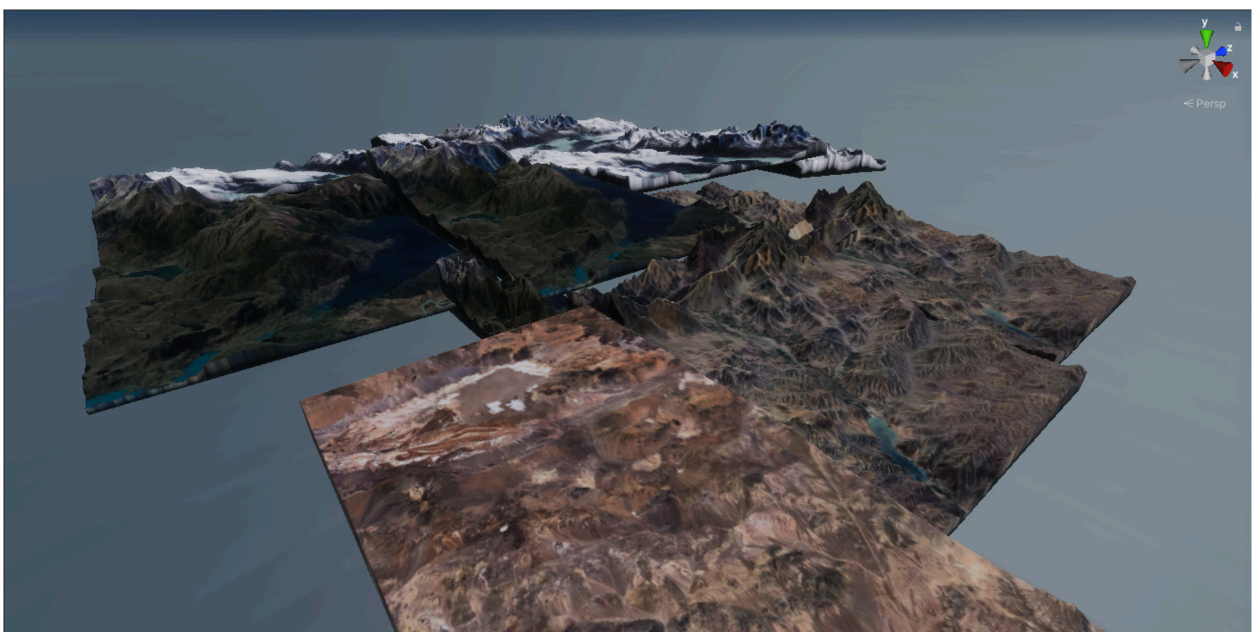


Figure 18. Satellite images rendered as 3D landscapes in a virtual environment.

[322] https://www.blender.org/ accessed January 23, 2025
[323] https://github.com/domlysz/BlenderGIS accessed on January 23, 2025
[324] https://www.openstreetmap.org/ accessed on January 23, 2025
[325] Found in the Unity asset store https://assetstore.unity.com/ accessed January 23, 2025

3.4.2.2 Mapping terrain features to musical parameters

A separate score was created as a guide for performers to interpret the 3D graphic score. Each landscape has a variety of musical parameters associated with it, with some of these parameters directly mapped to features of the terrain. Musical parameters such as character, pitch material, tempo, rhythm and dynamic create a space of musical possibilities for the performers. Meanwhile, for each landscape a specific terrain feature is directly linked or mapped to one or two musical parameters. Which feature is mapped to which musical parameter differs from one landscape to the another. The guide helps performers to understand what musical parameters they can explore freely while paying attention to a specific feature of the landscape to musically interpret. The musical parameters associated with each landscape are summarized in table 5.[326]

Landscape	Antofagasta (Desert)	Valle del Elqui (Mountain Valley)	Torres del Paine, (Deep Forest)	Torers del Paine, (Glaciar)
Character	Lonely	Playful	Dramatic	Mysterious
Pitch Material	E, G, C#, G#	D, F, A, C, E	Atonal/Chromatic	Emaj7 & F#maj
Tempo	Largo	Andante	Vivace	Adagio
Rhythm	Whole notes	Syncopated	Whole notes and 8th note runs	Arpeggio and sustained whole notes
Dynamic	pp to ff	n/a	p & f, contrasting	p
Terrain feature & Sound mapping	Color palette = Timbre	Perceived height = Tempo & articulation	Land/water mass = Rhythm & extended technique	Ice/rock mass = Register

Table 5. Summary of musical parameters associated with each landscape in *Senderos.*

[326] The complete score and performances of the piece can be found in the appendix section.

The selection of the locations came from a personal motivation. These are all places where I had a formative experience traveling with friends and family in Chile. I saw this composition as a way to remember these places while reflecting on the fragility and malleability of memory. To some extent, to remember is to internally travel through emotions, images, colors, sounds, scents, or any other object or sensation that will help us traverse through that memory space. On a conceptual level, the piece encapsulates the idea of memory as being a reinterpretation of the past, a reinterpretation that is artificial since it can never match the reality of the lived experience. Memory can (and will) continue to change as we remember again in the future. This is reflected in the visual representation of the landscape, an artificial map, recognizable from a distance, but ultimately a representation, intangible and imperfect. Because the music is not exactly predetermined, but rather a parametric space that musicians explore, it echoes the idea of memory as an exercise that is not always exactly the same. The act of remembering changes with each instantiation. In this sense, the choices for the musical parameters associated with each place is in itself my own exercise of remembering those places. It is an attempt to frame those memories within a musical parametric space for performers to explore.

3.4.3 *Paisajes Oníricos/Dreamscapes*: interactive 3D graphic score

Building upon the experience of composing a graphic score with 3D landscapes, *Paisajes Oníricos/Dreamscapes* incorporates interactive elements to control a 3D graphic score. It is a piece for open instrumentation and any number of performers. In the composition, a 'virtual conductor' moves through 3D space directing their view towards different graphical elements in the virtual environment. The composition reflects on the representation of graphic scores as drawings on a two-dimensional plane. The piece expands on this notion by allocating the graphical elements in 3D space, giving them mass and volume. By incorporating a spatio-temporal dimension, the virtual conductor can control the perspective and the pace of the composition via virtual navigation. The complete score can be found in the appendix of this document.

3.4.3.1 Visual design and 3D modeling

The virtual environment was designed using Blender. The design took inspiration in the work of late sixteenth century Dutch architect and painter Hans Vredeman De Vries. His work combined elements from different artistic currents of his time, including Gothic, Renaissance, Baroque and Mannerism; all relevant in the Netherlands in the late sixteenth century. Although there is no known building of his, many of his paintings are of architectural vistas.[327] His engravings in *Perspective* (figure 21), which demonstrate the rules of perspective and depict much of his architectural imagination, served as the main inspiration and reference for the design of the virtual environment.

The engravings suggested a minimalist design which could serve a twofold purpose: 1) diverge from the hyperrealism that is common to video games nowadays,[328] and 2) favor a more personalized and aestheticised design.[329]

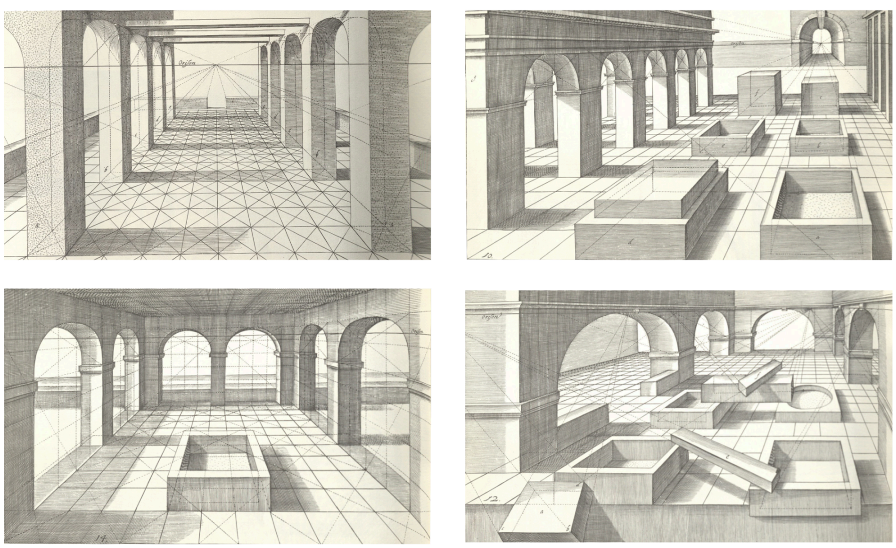


Figure 21. Engravings from *Perspective* by Hans Vredeman De Vries.

[327] Vredeman de Vries, Hans. *Perspective*. Dover Publications, 1968.
[328] https://www.nytimes.com/2024/12/26/arts/video-games-graphics-budgets.html accessed on January 17, 2025.
[329] Atherton, Jack, and Ge Wang. “Doing vs. Being: A philosophy of design for artful VR.” *Journal of New Music Research* 49.1 (2020): 35-59.

The virtual environment for the composition consists of three different sections, as shown in figures 22 and 23, each of these is associated with a musical movement:

1) A long narrow hall with beams spread along its path (figure 24).
2) Multiple corridors on the ground floor that contain various 3D objects (figure 25).
3) A maze on the second floor made from the ruins of buildings (figure 27).

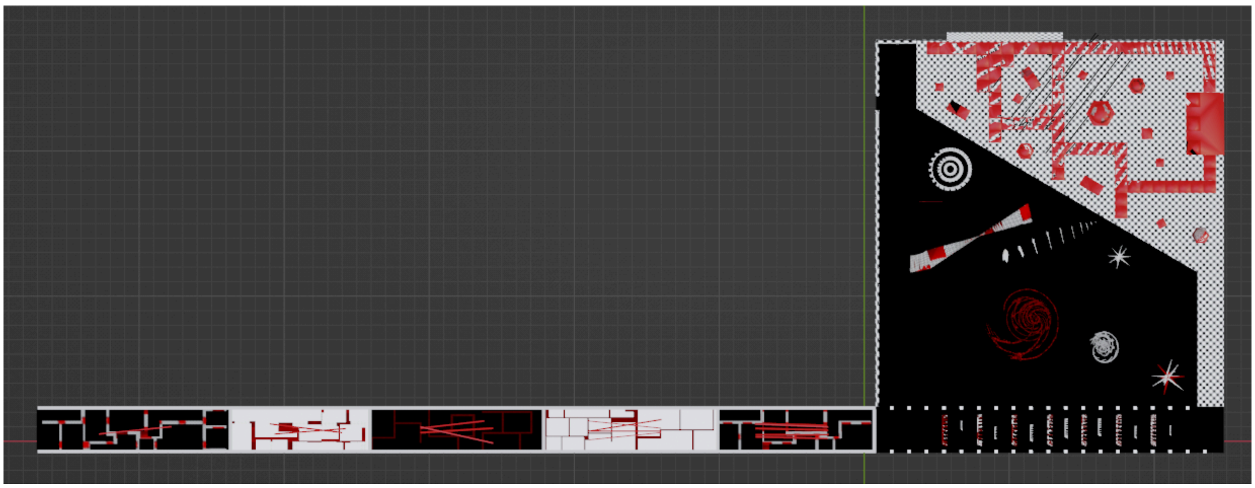


Figure 22. Top view of the virtual environment in Blender.

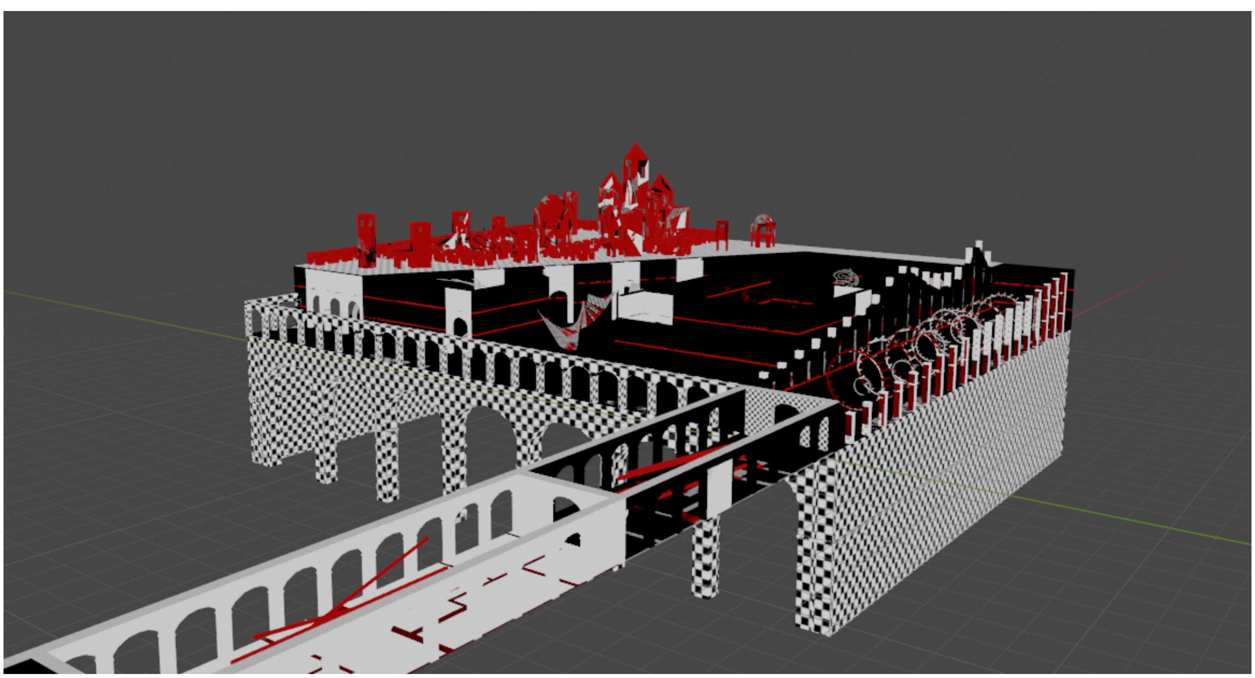


Figure 23. Side view of the virtual environment in Blender.

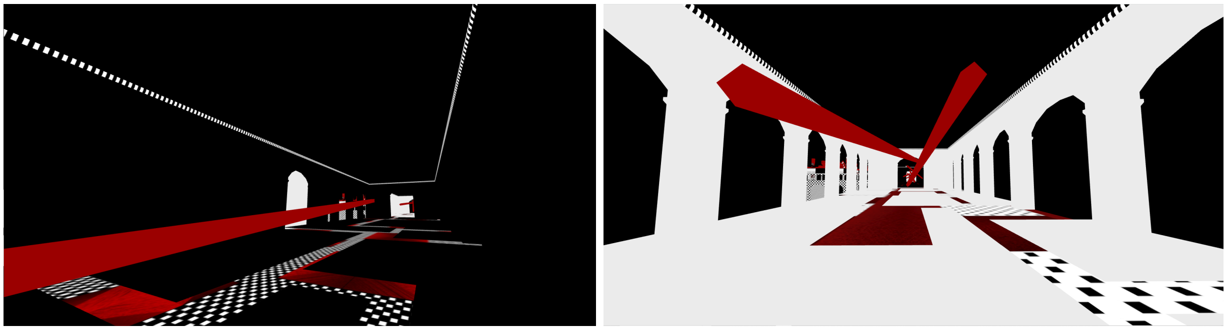


Figure 24. Snapshots from the conductor's first person perspective of the long narrow hall.

The 3D objects that inhabit the virtual environment were designed in Blender. Using different 'modifiers' in the software, complex procedures were applied on the arches and pillars from the first section to create many of the 'spiraling' objects for the second movement of the virtual score. Some of these examples are shown in figure 25.

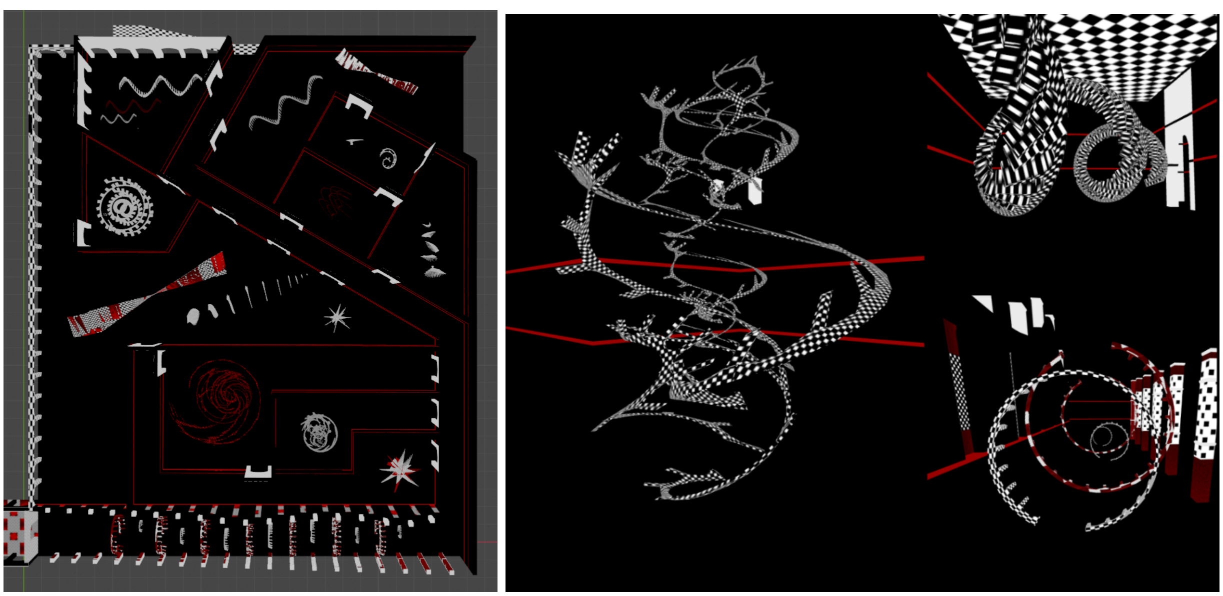


Figure 25. Top view of section 2 (left). 3D objects in the ground floor corridors (right).

The maze for the third movement of the composition was made from buildings designed using the web-based 3D modeling tool Tinkercad, as shown in figure 26.[330] The buildings were later imported into Blender and spatially rearranged. Applying the ‘boolean’ modifier tool, geometric figures were subtracted from the buildings to introduce ‘cuts’ and ‘openings’ in them.

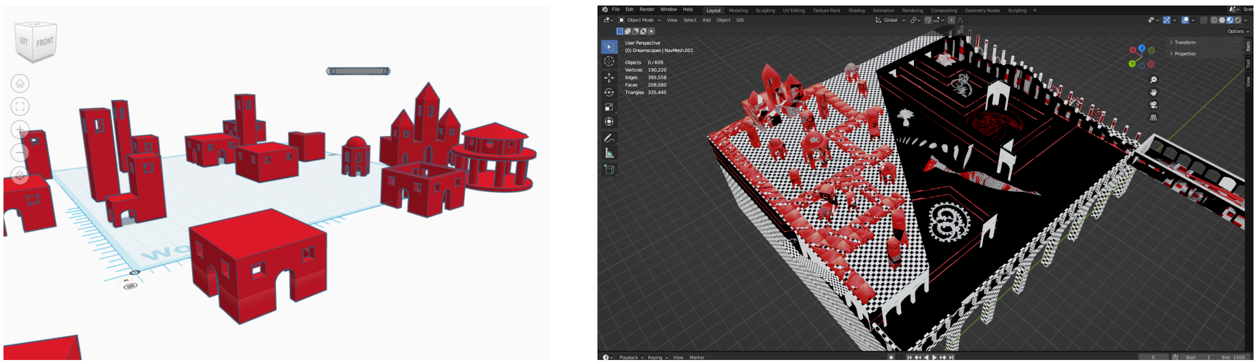


Figure 26. Models in Tinkercad (left). Top view of the third section’s maze (right).

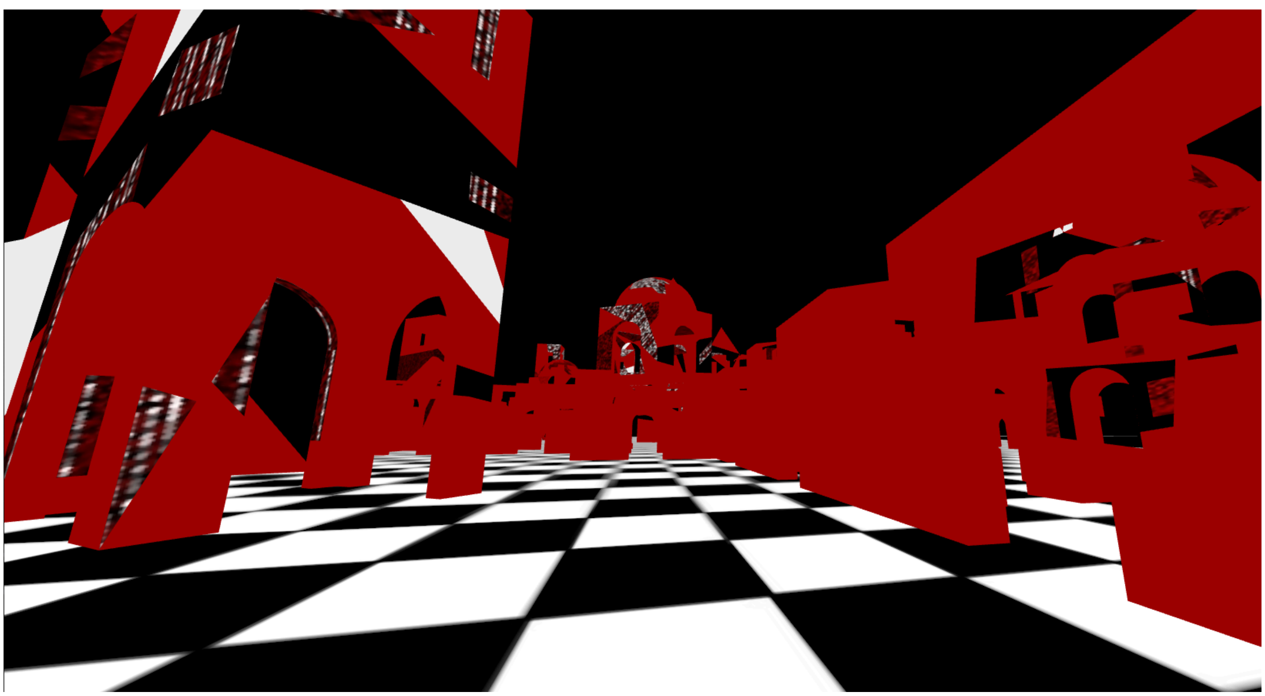


Figure 27. Conductor’s first person view from inside of the third section maze.

[330] https://www.tinkercad.com/ accessed on January 17, 2025.

3.4.3.2 Implementation

The first version of the virtual environment was implemented in Mozilla Hubs, a web-based virtual world platform that enabled users to create 3D environments and easily share them online for people to visit.[331] The Hubs-Blender exporter add-on[332] allowed for a streamlined export from Blender into Hubs, making it easy for users to design in Blender and directly export into the Hubs platform with multiple functional features such as virtual navigation, lighting, and animations, among other elements. This also made it easier for performers with no experience in XR to access the virtual environment by simply typing in the URL in a web-browser. After the shutdown of Mozilla Hubs, the virtual environment was ported to Unity with all the features and functionality the Hubs version had.

For a live performance, one of the performers or someone acting as a conductor has to run the application on a Windows machine and navigate the virtual environment. The view has to be projected on a screen for the audience to follow along with the musicians. Other setup configurations are also possible, like having a smaller screen on stage for the performers to view without having to turn their backs to the audience. The two setups are summarized in figure 28.

[331] It has shut down its operations since May 2024 and now the code is maintained by the Hubs foundation. https://hubsfoundation.org/ accessed on January 17, 2025.
[332] https://github.com/Hubs-Foundation/hubs-blender-exporter accessed January 17, 2025.

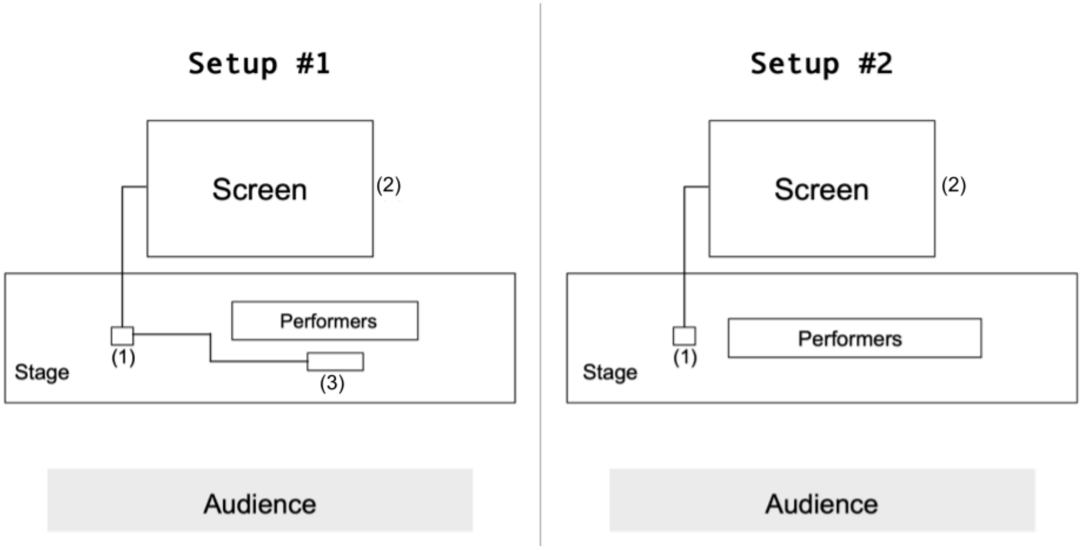


Figure 28. Setups for live performance: (1) computer running the virtual environment, (2) screen for the audience, (3) screen for performers on stage.

3.4.3.3 Virtual Conductor

The virtual conductor manages the virtual environment in real-time together with the rest of the performers. This requires someone to perform virtual navigation using a mouse to control the camera view and the keyboard in a 'WASD' configuration (commonly found in FPS-style video games) to move. The conductor does not perform any sounds, (just like music conductors in the real world), and they are in charge of advancing the score, controlling the pace of the piece. In this instance the virtual conductor also has agency over the structure of the piece as they can choose different pathways throughout movements #2 and #3. For example, there is a suggested route for the virtual conductor in the second movement, but they are encouraged to also experiment with different routes, resulting in the encounter of different graphical elements, and alterations to the duration of the piece. In the third movement, the virtual conductor is asked to convey a sense of urgency and chaos that has to be reflected in the erratic movements of the camera. These movements change the color palettes in the camera view, which in turn give cues to the musician to shift between frenetic loud noises and abrupt moments of silence. In this way, interactions in virtual environments enable the performance of non-musical media that has an impact on the sonic outcome of performers.

3.4.3.4 Mapping 3D graphical elements to music parameters

Different 3D objects are placed throughout the virtual environment. The performers interpret these objects using a separate guide that outlines mappings between the 3D objects and musical parameters, with specific instructions outlined for each section/movement of the composition.

Movement #1 - Red Beams

Throughout this long narrow hall, the virtual conductor encounters red beams in the direction of their movement. The red beams suggest note onsets, pitch, duration, glissando (if applicable) to the performers as shown in figure 29. The hall is divided in five sections, each with a different number of beams. The result is five instances of long sustained notes/chords, increasing in density on each repetition. A custom Max/MSP patch is given to the performers for them to record and playback each of the five instances. The loops created in this first movement establish a sonic identity that accompanies them throughout the second movement.

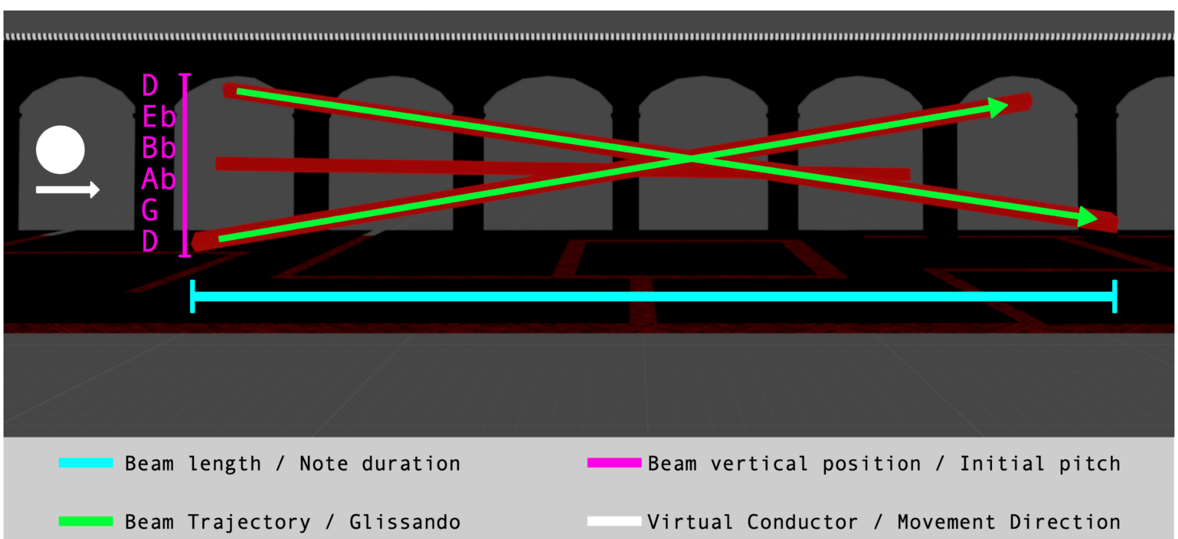


Figure 29. Mapping of music parameters to the red beams in movement #1.

Movement #2 - Vortex

The virtual conductor moves throughout the maze in the ground floor passing by various 3D objects, labeled as 'vortexes' given their swirling movement. Performers are asked to musically interpret the vortexes as they come in close range to them, as if the sound was emanating from the objects themselves. Performers are asked to create a sonic identity for each of the vortexes based on their color, shape, rotation speed, and volume, as shown in figure 30. Suggestions are indicated on how different parameters from the vortex can be mapped to tempo, articulation, extended techniques, and register.

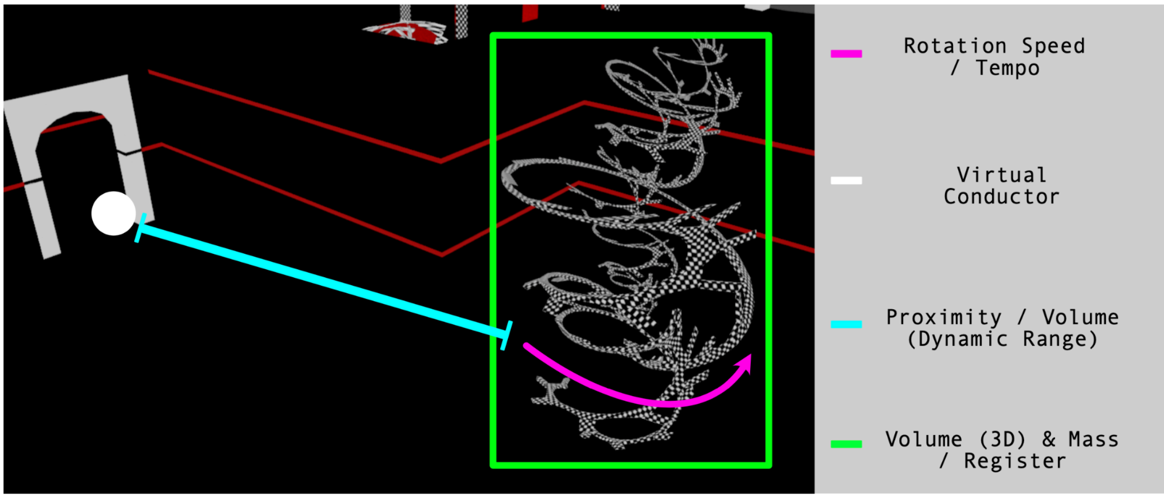


Figure 30. Mapping of music parameters to the vortex 3D objects in movement #2

Movement #3 - Maze

In the third section, the virtual conductor moves through a maze of ruined buildings as shown in figure 29. Performers are given more general instructions, and are asked to play frantically and chaotically. With the one caveat of doing abrupt pauses whenever the screen turns red. The virtual conductor is instructed to move frantically as well, encouraging colliding into the walls to turn the screen color red. As the virtual conductor exits the maze, the sounds from the performers fade out.[333]

Figure 31. Conductor's first person view of the maze for movement #3.

[333] The complete score and performances of the piece can be found in the appendix section.

3.5 Immersive storytelling: multi-channel soundscape composition

This section describes the ideation and composition process for *Funeral For A Whale*, a multi-channel soundscape composition. The first ideas for the piece emerged from initial experimentations with recording and processing sounds from servo motors of a robotic display used for telematic music performance. Initially, the intention was to create a soundscape composition that would imagine the soundscapes of the future based on different mechanical and robotic sounds. However, processing the servo motor sounds through time stretching and pitch shifting a few octaves down resulted in a sound that suggested a 'big creature moving underwater'. Although this music composition does not involve a visual nor interactive virtual environment, immersive multi-channel sound systems could be considered as auditory displays for virtual environments. In this case, appealing to narrative form that involves a participatory listening from the perspective of the audience.

3.5.1 Animal grief, soundscape composition, and participatory listening

A series of news articles have been published describing expressions of grief carried out by marine mammals.[334] For example, an orca mother carrying their dead calf for days, mourning their deceased offspring. In *How Animals Grieve*, anthropologist Barbara J. King writes about the anecdotes of marine researchers on seeing dolphins engage in similar behaviour for an extended period of days. In some instances other dolphins would escort the mother and the dead calf, defending the corpse from seagulls and at some points helping the mother by supporting the corpse with their own bodies.[335] Emotional responses of this kind have also been observed in whales, and researchers hypothesize that it would explain some of the mass strandings that occur in various whale species as part of an 'emotional

[334] https://www.nationalgeographic.com/animals/article/whales-death-grief-animals-science
https://www.npr.org/2018/07/31/634316124/grieving-mother-orca-carries-dead-calf-for-more-than-a-week-over-hundreds-of-mil
https://www.cbc.ca/radio/asithappens/as-it-happens-tuesday-edition-1.4768344/orcas-now-taking-turns-floating-dead-calf-in-apparent-mourning-ritual-1.4768349 accesed January 19, 2025.

[335] King, Barbara J. *How animals grieve*. University of Chicago Press, 2013.

contagion', where a single whale becomes stranded due to distress or injury, while others follow to not leave their kin behind, becoming stranded themselves.[336]

In composing *Funeral For A Whale* the intention was to imagine a narrative structure that would explore that space of grief and mourning, both from psychological and physical perspectives. For references I listened to different soundscape composition pieces by Luc Ferrari, Maggie Payne, Barry Truax, and Hildegard Westerkamp, and looked at the different techniques and structures used by many of the composers that pioneered soundscape composition.[337] Truax describes the different approaches to soundscape composition as falling between 'found sound' and 'abstracted sound.' The first one refers to sounds from the real-world that retain a certain degree of recognisability, invoking associations in the listener. The second refers to the processing of sounds that oftentimes abstracts them from their real-world origins through techniques such as time stretching and transposition, among others.

Katharine Norman argues that tape music that employs sounds from the real world as musical material is enriched and formed through the listener's imaginative response. She talks about the different ways of listening in the context of tape music, where the synthesis between 'referential' and 'reflective' listening, activates both memory and imagination in the user depending on how the real world sonic materials are organized and interact with abstract musical sounds. She argues that listeners can contribute creatively to the music through active imaginative engagement.[338]

3.5.3 *Funeral For A Whale*: Structure and Description

Truax identifies the 'moving perspective' as a common approach used in soundscape composition as it can depict and suggest narrative structures through the illusion of a listener that moves through different acoustic environments.[339] I used this approach to structure the piece in a way that suggests a narrative in the form of a journey, moving from the deep sea water to the surface and beyond. One of the goals with

---

[336] King, Barbara J. *How animals grieve*. University of Chicago Press, 2013.
[337] Truax, Barry. "Genres and techniques of soundscape composition as developed at Simon Fraser University." *Organised sound* 7.1 (2002): 5-14.
[338] Norman, Katharine. "Real-world music as composed listening." *Contemporary Music Review* 15.1-2 (1996): 1-27.
[339] Truax, (2002).

this piece was to balance 'referential' and 'reflective' listening for the audience to actively imagine the narrative and participate in its interpretation.[340] The composition is roughly divided in three sections with the following time frames:

Section I. 00:00 - 05:00
Section II. 05:00 - 09:30
Section III. 09:30 - 12:50

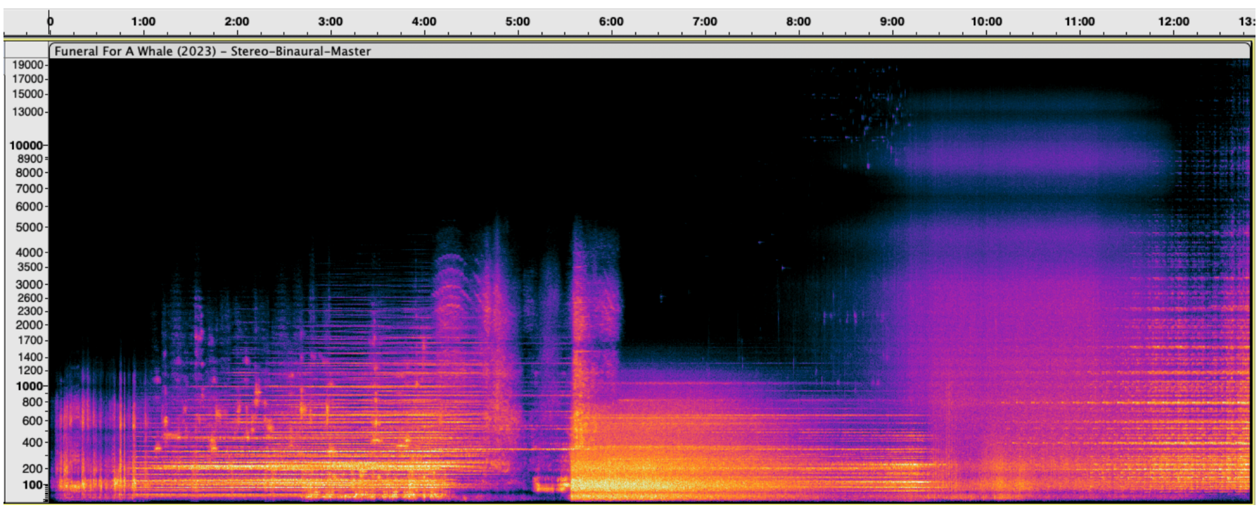


Figure 32. Spectrogram of *Funeral For A Whale.*

These sections are not clear cuts between one another as the piece makes use of long crossfades between multiple layers of sound to smoothly transition between each section. In the following paragraphs the materials and techniques used in each section are described, together with how they were applied to transit from one section to the next while creating moments of tension and release throughout the piece. In describing the piece, I will most likely also be giving away my own interpretation of the piece, and how I imagine what each of the sounds represent in this narrative structure.

[340] Norman, Katharine. "Real-world music as composed listening." *Contemporary Music Review* 15.1-2 (1996): 1-27.

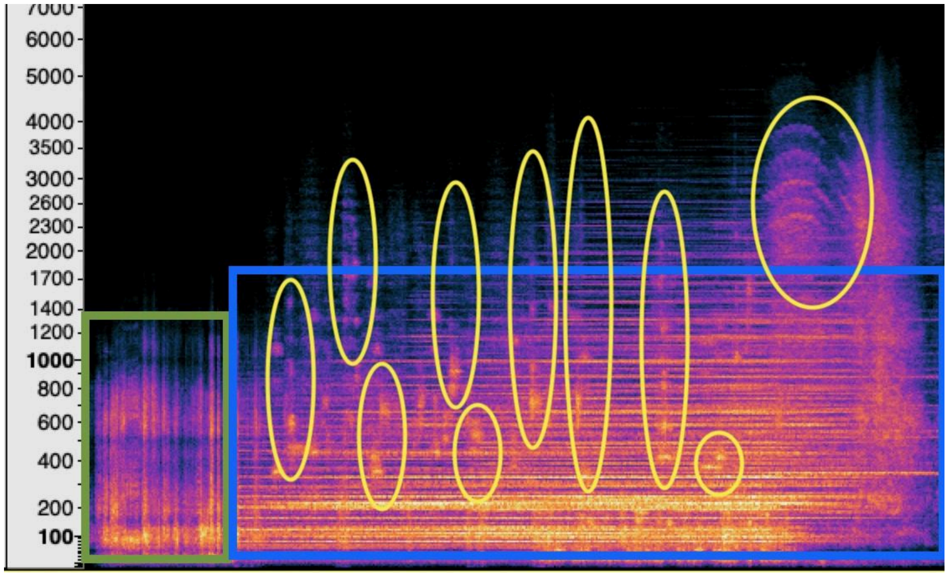


Figure 33. Spectrogram of section I of *Funeral For A Whale*.

3.5.3.1 Section I, Underwater

The piece begins by playing back a sound that suggests a large mass moving in an underwater environment, as depicted inside the green box in the spectrogram. This sound is eventually brought back halfway through the section As the sound circles around the listener, a drone on the bass register comes in, followed by mid to high range hisses that suggest whales singing. Drones are continuously added in the low mid and high registers (blue box, figure 33), rotating at a slow pace and at different rates around the listener. The drones are layered together in a harmony that continues to increase in dissonance and intensity throughout this section. Inside the blue box it is possible to see where new drones are added while increasing in intensity towards the end of the section. The yellow ovals depict the 'hissing', 'breathing', and 'whale song' sounds that are coming in and out throughout the section, spatialized as coming from different angles to suggest creatures swimming around. Towards the end of this section the drones in the mid and lower register fade out as a 'hissing' and 'airy' sound takes over into the beginning of section II.

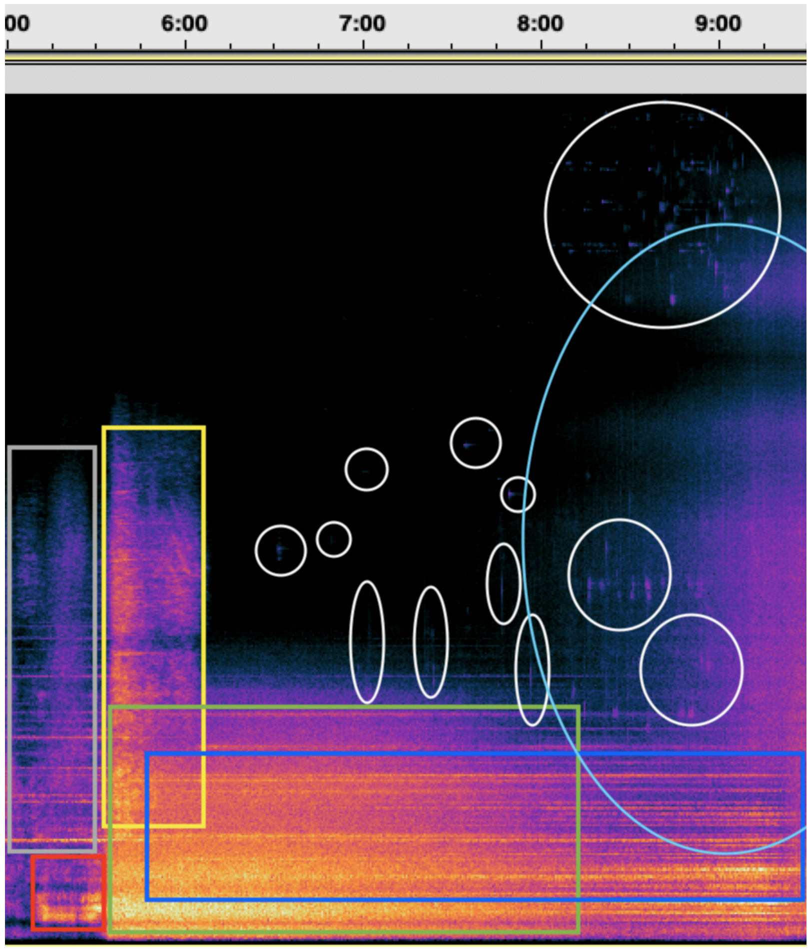


Figure 34. Spectrogram of section II of *Funeral For A Whale*.

3.5.3.2 Section II, Decay & Corpse Surfacing

The second section begins with the continuation of the ‘hissing’ and ‘airy’ sounds (grey box, figure 34) at a much lower intensity releasing much of the tension that had been built up from the previous section. With the exception of one continuous drone in the mid-high register maintaining a subtle dissonance. As the hissing fades out a low sound (red box) that resembles the one at the beginning of the piece anticipates the burst of sound that is to follow. A sudden loud ‘hissing’ sound (yellow box) and low end ‘water stream’ (green box) come into play at a high intensity. By spatializing the water stream on the left and right sides of the listener and applying a movement from front to back, the water stream gives the sensation of rapid movement underwater. The hissing completely fades out while the water stream continues, and new drones come in, in the mid register. As the water streams and drones settle, high pitched sounds start to appear in the distance, with metallic and insect-like qualities (white ovals). As the water stream fades out, the drones become more prominent, and the high-pitched sounds start to blend in with sounds of water droplets (blue oval). The high-pitched sounds accumulate in density together with a very prominent dissonance from the layered of drones, these gently fade out, into the increasingly enveloping sound of rain and into the third section.

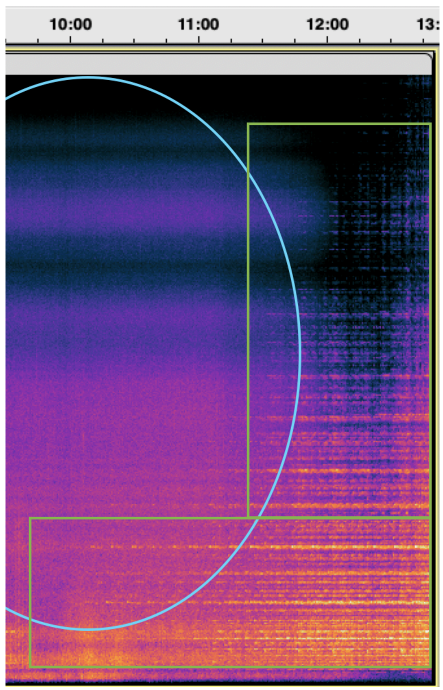


Figure 35. Spectrogram of section III of *Funeral For A Whale*.

3.5.3.3 Section III, Procession and Ascension

The third section begins with a field recording of rain, thunder can be heard in the distance with some near droplets as well, suggesting that the listener has emerged from underwater. The rain functions as another moment of release after the accumulated tension at the end of the second section (blue oval, figure 35). As the rain becomes more intense, melodic sounds of brass start to come into the scene (green box), suggesting a procession. First low and mid range brass sounds, some of them are transposed down enough to suggest and resemble some of the 'whale song' sounds from the first section. There is somewhat a recognizable pattern in the brass sound, but no melodic motif. As the brass sounds become clearer, with sounds in low, mid and high registers, that continue to interject each other with no discernible pattern, the rain starts to fade out. As the rain is left behind, the listener is enveloped by the sounds of brass, which start swirling around the listener growing in intensity, suggesting an 'ascension'. Higher pitched brass sounds come in within the last thirty seconds and in the last ten seconds, the sounds swirl at an increasingly faster pace, volume and intensity builds up steadily and rapidly only for the sound to be cut off suddenly.

3.5.4 Spatial audio tools, sound materials and audio processing

Funeral For A Whale was composed in the Jefferson Starship Studio using the open source plug-in suite from Envelop[341] for Ableton Live and the 8-channel Genelec speaker configuration for playback. Other tools included Audacity (Paulstretch) for the time stretching of multiple samples, the Max for Live Convolution Reverb, and the RX 6 plug-in suite from Izotope.

For the piece a variety of sound materials were used, including: field recordings of rain, water streams, servo motors, samples from instruments and ensembles including piano, brass, symphonic orchestra, and sounds generated with the ARP 2500.

3.5.4.1 Time stretching/compression and transposition

Many of the materials used in the composition were processed using time stretching/compression. For example, the opening sound in the piece that suggests a large mass moving underwater was made from a recording of multiple functioning servo motors. The servo motor recording was transposed down somewhere between one and two octaves in Ableton Live. By turning the 'warp' function off in the DAW, this results in the recording being stretched proportionally to the transposition.

The 'hissing', 'breathing' and 'whale song' sounds were also made from the recordings of servo motors. These were first stretched in Audacity using the Paulstretch algorithm, eliminating the mechanical

[341] https://envelop.us/page/software accessed January 20, 2025.

clicking sound of the motors, giving it a more continuous quality. These stretched materials were later further edited into single gestures with fade ins and fade outs, and transposed above and below their original pitch without warping in Ableton Live. The drones in sections I and II were made using the same workflow, time stretching in Audacity, transposition without warping in Ableton Live, but these were made from short orchestral and piano samples.

3.5.4.2 Cleaning sound samples

The RX-6 Spectral De-noise was used extensively to process the different materials used in the piece. Its application went beyond the usual 'cleaning' of sound samples and was used to purposely make the sounds for sections I and II mimic the sensation of being underwater, as it can be seen in the spectrograms for these two sections (figure 30). For example, the sounds of the water stream in section II, were made from recordings of an open bathtub faucet and by hand gestures on a bathtub filled with water. These were dramatically altered using the denoiser as an equalizer to get rid of the mid high and high end of the recording. This was also applied on the drones in this section to match the spectral content of the water stream making it easier to layer together and transition from the water stream to the drones in section II.

3.5.4.3 Layering and spatialization

The composition makes extensive use of layering paired with the spatialized movement of the layered sounds. Almost all of the identifiable elements in each section is made up of several layers or tracks, with minute variations on the way the sounds are processed.

For example, the brass sounds in section III were made from brass samples that were loaded onto an Elektron Digitakt. A four bar loop was created using the sequencer where they were transposed and warped in duration. Multiple recordings of the same loop were made in a DAW, with minute variations in editing to make the loop less recognizable. These loops were layered together, and spatialized as if a 'whale brass band' was coming from a particular direction to eventually envelop the listener.

In another example, the ‘hissing’ and ‘whale song’ sounds in the first section, consist of multiple layers, spatialized as moving together or in opposition. The drones in the first section are made from the same sample, with differences in transposition (and thus duration), and spatialized as moving around the listener. The same is applied for the ‘water stream’ sounds, which are made of multiple layers of the same recording, with differences in EQ and transposition.

The original stereo recording of rain received a similar treatment to achieve an enveloping sensation on an octophonic system and simulate being in a rainy environment. Multiple edits from the same recording were layered together, with transpositions ranging from a fifth higher relative to the original pitch and an octave below. The tracks were equalized and denoised differently, with the goal to cover most of the hearing spectrum with multiple sources to create a sonic environment that felt like natural rain.

3.6 3D Corpus Navigator: a virtual musical instrument for sound corpus manipulation[342]

This section describes the design and implementation of the 3D Corpus Navigator, a FPS-style interface made in Unreal Engine 5 that relies on virtual navigation to control Corpus Based Concatenative Synthesis (CBCS) in Max/MSP using the FluCoMa library for analysis and playback. In this virtual musical instrument, the user is turned into a virtual 'playback head' that can fly in 3D space. The user can interact with a 3D point cloud by flying through the different points, activating the sound sample associated with each point upon contact. The user can dynamically switch between different speed settings, using their position and velocity data to control different mappings of sound effects that extend the sonic gestures made with the sound corpus in Max/MSP.

As outlined in Chapter 2, multiple artists and scholars have developed virtual musical instruments that map virtual navigation in virtual environments to different musical parameters, such as applying virtual navigation as a procedural system to control parameters for sound synthesis in real-time or as an audiovisual sequencer. Many have explored the use of gamified interfaces that resemble those found in popular video games genres such as the 'First Person Shooter' (FPS). Hamilton modified the game engine of a FPS to turn the game mechanics into sonic interactions for a multi-user electro-acoustic music performance.[343] Berthaut et al. created a collaborative multi-process instrument which relies on the interaction techniques used in FPS video games.[344]

---

[342] Portions of this section were originally published in Vilaplana Stark, M. A. “Developing a 3D interface for sound corpus manipulation in virtual environments,” *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Saint Malo, France, 2025, pp. 643-646 © 2025 IEEE

[343] Hamilton, Robert. “Maps and legends: Designing fps-based interfaces for multi-user composition, improvisation and immersive performance.” *International Symposium on Computer Music Modeling and Retrieval*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[344] Berthaut, Florent, et al. “First person shooters as collaborative multiprocess instruments.” *New Interfaces for Musical Expression*. 2011.

3.6.1 Interactive systems that use Corpus Based Concatenative Synthesis

Corpus Based Concatenative Synthesis (CBCS) is a technique that has been widely used in computer and electro-acoustic music.[345] [346] [347] Multiple interfaces and gestural controllers have been developed for real-time interactions with CBCS in musical performance. [348] [349] [350] 3D interactive systems have also been created to control corpus-based concatenative synthesis. In *Plumage*, sound grains are spread in space in the form of virtual feathers. It uses controllable virtual playheads that can move in cyclic trajectories. The playheads have associated 'triggers' that orbit around the playheads and activate the sound grains as they come in contact with the 'feathers'.[351]

Immersive systems to explore musical interactions with CBCS have also been developed using VR and AR systems. Zappi et al. developed an immersive installation using inertial and marker-based tracking systems where users are free to use their hands to interact with a sound corpus distributed in virtual space.[352] Neupert created a VR system using a low-cost phone based setup with a gesture tracker to immerse users in a virtual environment where they can interact with a 3D point cloud and perform concatenative synthesis through theremin inspired gestures.[353] Halac and Addy developed an interactive VR experience that displays timbre descriptor data from a sound file as a point cloud. Using handheld-controllers users can create musical sequences and patterns using the samples from the 3D point cloud.[354] More recently, Berthaut developed a mixed reality musical instrument using a Spatial

[345] Hackbarth, Benjamin, et al. "Composing morphology: Concatenative synthesis as an intuitive medium for prescribing sound in time." *Contemporary Music Review* 32.1 (2013): 49-59.
[346] Schwarz, Diemo, et al. "Real-time corpus-based concatenative synthesis with catart." *9th International Conference on Digital Audio Effects (DAFx)*. 2006.
[347] Stine, Eli. "Creating Immersive Electronic Music from the Sonic Activity of Environmental Soundscapes." *Iui workshops*. 2019.
[348] Beller, Grégory, and Georges Aperghis. "Gestural control of real-time concatenative synthesis in luna park." *P3S (Performative Speech and Singing Synthesis)* (2011).
[349] Schwarz, Diemo, et al. "Musical applications of real-time corpus-based concatenative synthesis." *International Computer Music Conference (ICMC)*. 2007.
[350] Zbyszyński, Michael, et al. "Gesture-timbre space: Multidimensional feature mapping using machine learning and concatenative synthesis." *International Symposium on Computer Music Multidisciplinary Research*. Cham: Springer International Publishing, 2019.
[351] Jacquemin, Christian, et al. "Plumage: Design d'une interface 3D pour le parcours d'echantillons sonores granularises." *Proceedings of the 19th Conference on l'Interaction Homme-Machine*. 2007.
[352] Zappi, Victor, et al. "Concatenative synthesis unit navigation and dynamic rearrangement in vrgrains." *Proceedings of the sound & music computing conference*. 2012.
[353] Neupert, Max. *Exploring Concatenative Synthesis Units in VR*. Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2017.
[354] Halac, Fede Camara, and Shadrick Addy. "PathoSonic: Performing Sound In Virtual Reality Feature Space." *NIME*. 2020.

Augmented-Reality display to project 3D textures in physical space where users can control granular synthesis via hand gestures.[355]

3.6.2 Implementation

The integration of Unreal Engine and Max/MSP via OSC supports a wide range of creative possibilities for interactive computer music performance and composition. Unreal Engine 5 offers a robust interactive platform with enhanced graphics, while the FluCoMa library in Max/MSP offers a variety of tools for descriptor analysis, interactive sample playback, and live-processing given its configurability and scalability.[356]

3.6.3 Interface for descriptor analysis and general settings

The interface is a front end Max/MSP patch that makes it simple to work with any sound corpus owned by the user. Users can segment and analyze the sound corpus by using the interface to select from the different algorithms the FluCoMa library makes available and fine-tune its parameters, which can be saved and recalled with the preset object. These parameters give flexibility to the user on how to segment the sound corpus and analyse it, determining the spatial arrangement of the 3D point cloud in the virtual environment. Figure 36 shows the Max/MSP interface.

---

[355] Berthaut, Florent. “Musical Exploration of Volumetric Textures in Mixed and Virtual Reality.” In Proceedings of the International Conference on New Interfaces for Musical Expression, 2021.
[356] Tremblay, Pierre Alexandre, Gerard Roma, and Owen Green. “Enabling programmatic data mining as musicking: the fluid corpus manipulation toolkit.” *Computer Music Journal* 45.2 (2021): 9-23.

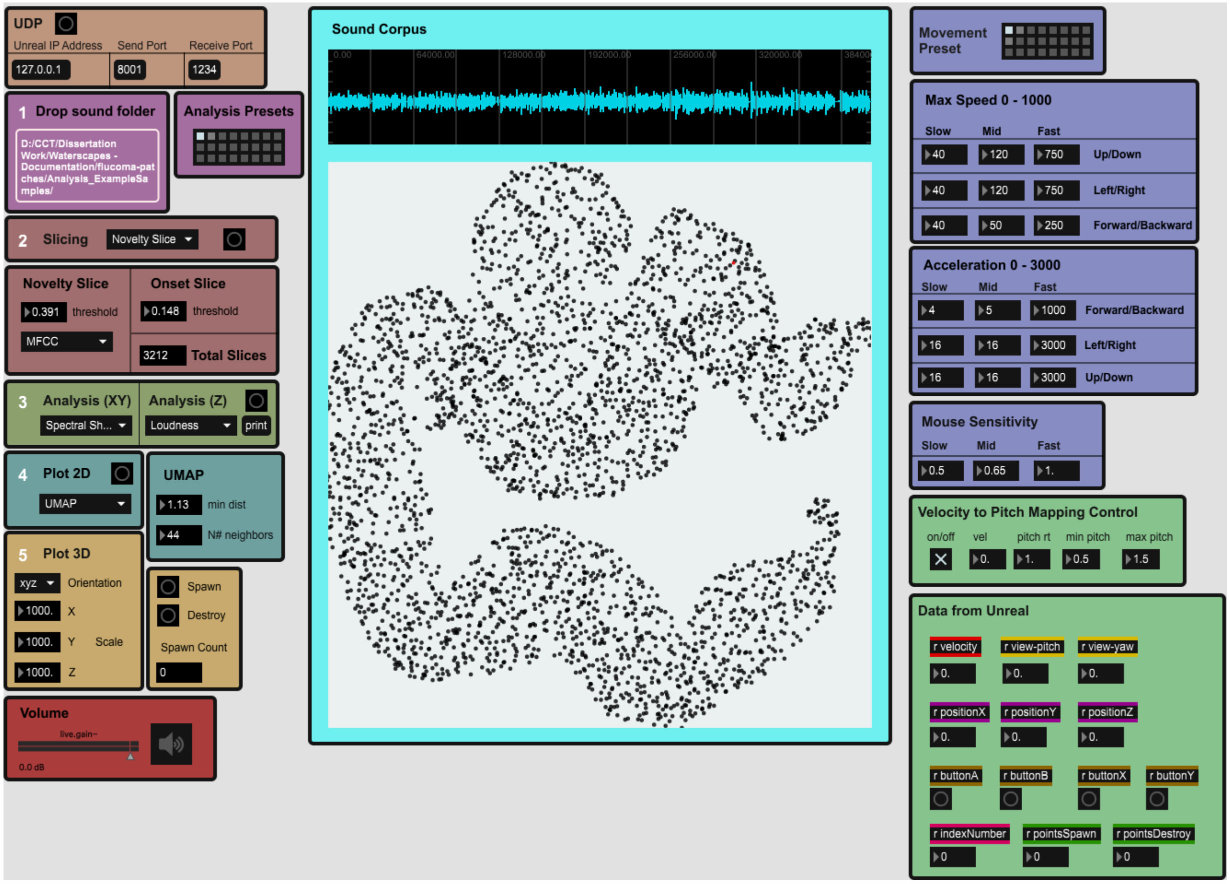


Figure 36. 3D Corpus Navigator Max/MSP interface. © 2025 IEEE

3.6.3.1 Segmentation, Descriptor Analysis, 2D and 3D plotting

The user can select between different onset detection algorithms to segment the corpus into sound slices of different durations. The slices (or samples) are then analysed to determine the arrangement of samples in the XY plane and the Z axis. For the analysis on the XY plane users can choose between different descriptor algorithms including: MFCC, Spectral Shape and MelBands. For the Z axis they can choose either to apply Loudness, Pitch, or Chroma analysis.[357] All the datasets generated are normalized in this step, before the data of each sample is sent to the virtual environment. The XY plane is shown in a 2D representation for the user to have a previsualization before plotting the sound corpus in 3D space. In the Max/MSP patch the user can adjust the orientation of the XY plane + Z axis and determine the scale of

[357] More information on the analysis algorithms can be found in their reference page https://learn.flucoma.org/reference/

each axis separately. The orientation options include all possible permutations between the three axes, indicated in the form: XYZ, ZXY, YXZ, etc.

3.6.3.2 Movement and playback settings

In this section of the patch the user can adjust the maximum speed and acceleration parameters for the three different speed settings. This gives the user flexibility to adjust the ranges of velocity and acceleration to something that feels comfortable for them, or be set to explore more challenging speed configurations. For the mouse and keyboard interface, the user can also adjust the mouse sensitivity associated with each speed setting.

The user has the option to play the samples at their original pitch, or to map the playback speed to the real-time velocity of the virtual playhead in the virtual environment. If mapped to the speed of the virtual playhead, the user can adjust the pitch range to which the velocity is scaled to i.e. two octaves above and below the original pitch. The implementation is polyphonic in order to trigger multiple samples simultaneously or with overlap.

3.6.4 Virtual environment and controls

Unreal Engine 5 has a user-friendly system for managing input mappings, and has an integrated OSC plug-in, allowing for communication with other music softwares like Max/MSP.

The user has a first-person view in the virtual environment, with no virtual embodiment whatsoever. The user starts in a seemingly empty space, a cube with disconnected vertices is placed in front of the user as seen in figure 37.

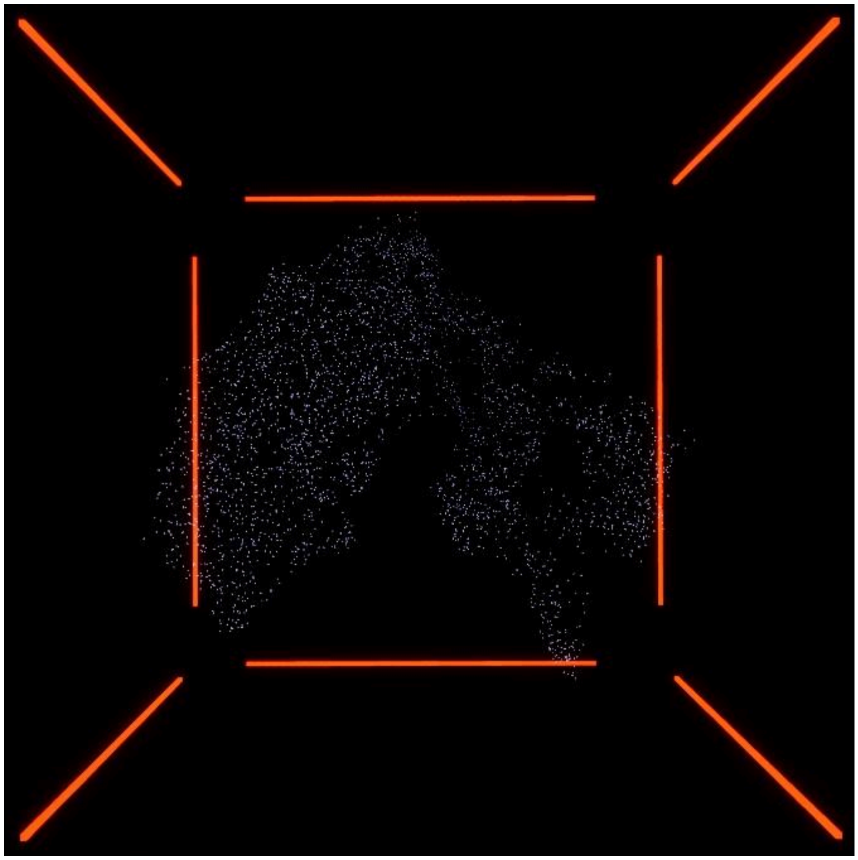


Figure 37. 3D point cloud in the 3D Corpus Navigator virtual environment.

3.6.4.1 Movement and interaction controls

The ‘WASD’ keyboard can be used to move front/back and left/right, plus the 'Q' and 'E' keys to move up and down respectively. The mouse is used to look around, which changes the direction of the forward vector. The user moves by default in speed setting 1, they can hold down the ‘shift’ key while moving to engage speed setting 2, and simultaneously hold down the ‘shift’ + ‘space’ key to engage speed setting 3.

It can also be controlled with a game-pad as shown in figure 38. This example is modeled after a Nintendo Switch Pro Controller, but any generic game-pad, such as an Xbox or PlayStation controller would also apply.

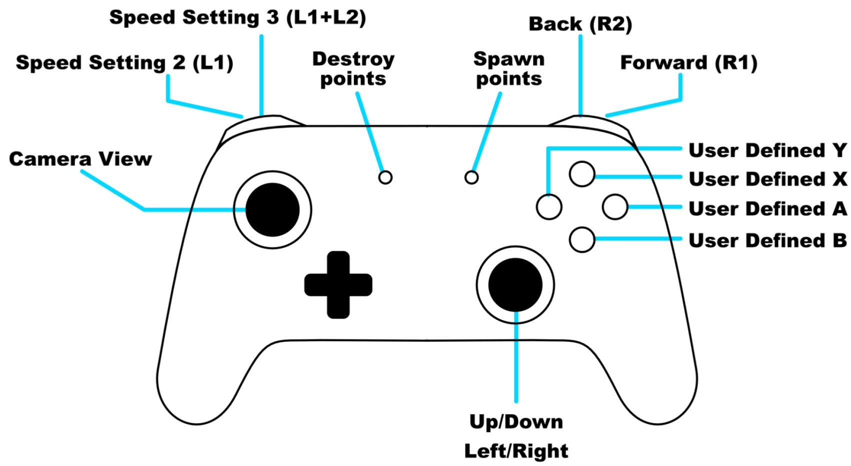


Figure 38. Game-pad controller input mapping for the 3D Corpus Navigator. © 2025 IEEE

The left joystick is used to look around, while the right joystick is used for up/down and left/right movement. Forward and backwards movement is controlled with the right hand triggers R1 and R2 respectively. The left hand triggers are used to access the different speed settings: trigger L1 for setting 2, and trigger L1 + trigger L2 for setting 3. The 'user defined buttons' can be used to extend the sonic interactions (discussed in the following section).

The user can press the 'enter' key or the 'start' button on the game-pad to spawn the samples as virtual 3D points. The points are spawned at a rate of ten milliseconds, with the option to play back their assigned sample in the process. This option can serve as a way to audibly survey the complete sound corpus while visualizing the process of populating the 3D space. Pressing the 'backspace' key or the 'select' button on a game-pad causes all points to be destroyed without any sample being triggered.

3.6.4.2 Sonic interaction

The user can interact with the 3D points via proximity, by passing through the points as if the user was a flying virtual playhead. The collision with 3D points is set as a 'trigger', meaning that there are no physically simulated collisions. As the user clips through the points to trigger sound samples, the velocity registered at the moment of contact can be mapped to determine the playback speed of the sample. Users can create new sonic material through the real-time reconfiguration of a sound corpus based on their movement trajectories and speed. The color of the 3D points changes when triggered. This serves to give visual feedback of which point has been activated while leaving a visual trail of the movement enacted by the user. Although the system might lack the precision to select sound samples when compared to other cursor based CBCS systems, the use of a FPS-style interface for interaction adds a degree of effort. Adriana Sá discusses the role of effort in sonic expression in audiovisual performances:

> "To threshold the performer's control over the instrument, and the unpredictability of sonic outcomes - so that the instrument affords sonic complexity, in a way that suits the performer's idiosyncratic expression"[358]

By making the instrument more 'difficult' to control for the users, it forces users into unpredictable outcomes in the process of learning and mastering the instrument. Whether how much 'effort' the instrument requires from the user in quantifiable terms is out of the scope of this project.

3.6.5 Composing and Performing with the 3D Corpus Navigator

This section describes the practical application of the 3D Corpus Navigator for the live music performance of *Point Cloud*, at the Technosonics festival in Charlottesville, VA.

3.6.5.1 Curating a sound corpus

The sound corpus is almost entirely responsible for the sonic identity of the piece, the curation of it is a fundamental part of the composition itself. The sound corpus used in *Point Cloud* comes from a variety of sound sources, including a sample pack of ARP 2500 sounds, vocal sounds, recordings from a nylon

[358] Sá, Adriana. "A method for the analysis of sound art and audio-visual performance." *Audiovisual e Industrias Criativas: Presente e Futuro* (2021): 575-589.

string guitar, an acoustic piano, various synthesizers, and sounds generated experimenting with Google's MusicLM.[359]

3.6.5.2 Sonic extension, audio effects and mapping

The 3D Corpus Navigator allows users to personalize and extend the sonic interactions beyond the real-time reconfiguration of sound samples via virtual navigation. The Max/MSP patch gives users access to incoming data from the virtual environment, including position, velocity, camera angle, sample index, and controller input. This enables users to integrate their own audio processing effects in Max/MSP and mapping them directly to the interactions in the virtual environment to create idiosyncratic configurations of the instrument.

Custom audio processing effects made in Max/MSP were mapped to the virtual interactions in the 3D Corpus Navigator virtual environment to perform *Point Cloud*. The four different mappings can be selected from the 'user defined buttons' dynamically, adding a layer of complexity to the sonic out by enabling the user to also concatenate different sound effects as they fly through the 3D point cloud. The four custom effects described in the following paragraphs also provide a sonic identity to the piece as part of the composition itself.

[359] Agostinelli, Andrea, et al. "Musiclm: Generating music from text." *arXiv preprint arXiv:2301.11325* (2023).

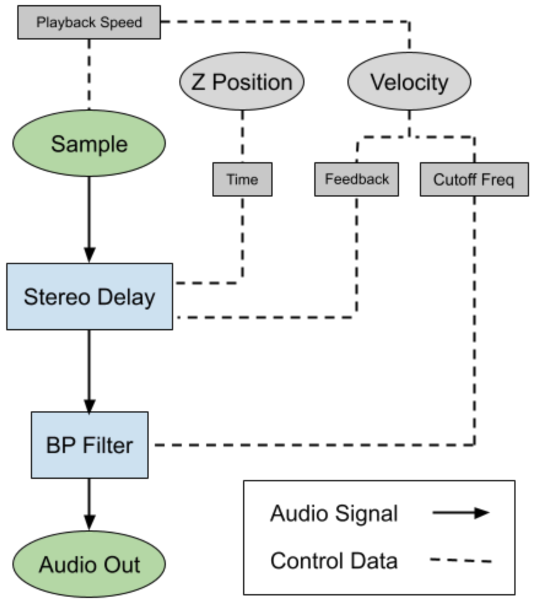


Figure 39. Mapping #1, stereo delay + bandpass filter.

Mapping #1: Stereo Delay + EQ

In this mapping the velocity magnitude of the playback head is mapped to the playback speed of the sample, the feedback of a stereo delay and the cutoff frequency of a bandpass filter, with relationship between velocity and the musical parameters being directly proportional. The position in z-axis (height) of the virtual playhead in the virtual environment determines the time of the stereo delay, in this case with higher altitude linked to longer delay times.

Mapping #2: Comb Filter + Ring Modulator

In this mapping the velocity magnitude of the playback head is mapped to the playback speed of the sample. The relatively short delay times of the comb filter are directly proportional to the height position (z-axis) of the virtual playhead, while the feedback depends on both the height position and velocity of the user. Velocity is also directly proportional to the modulation frequency of the ring modulator.

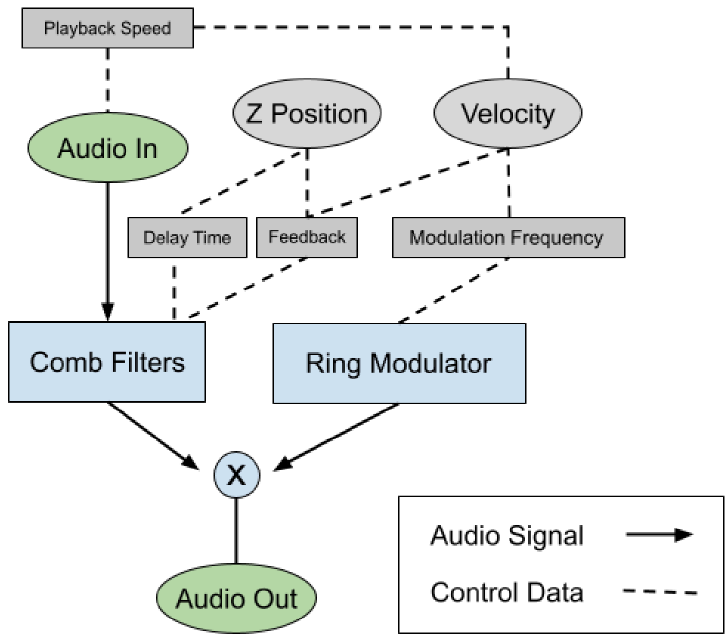


Figure 40. Mapping #2, comb filters + ring modulators.

Mapping #3: Harmonizer + Stereo Delay

In this mapping the playback speed is directly proportional to the z-axis position of the virtual playhead. Upon selecting this mapping, it randomizes the pitch ratio of the three voices of the harmonizer. The velocity is mapped to both time and feedback of a stereo delay in a directly proportional relationship.

Mapping #4: Harmonizer + Comb Filter

In this mapping the playback speed and comb filter delay time are directly proportional to the z-axis position of the virtual playhead. The velocity is directly proportional to the comb filter feedback and the pitch ratio of each of the three voices of the harmonizer.

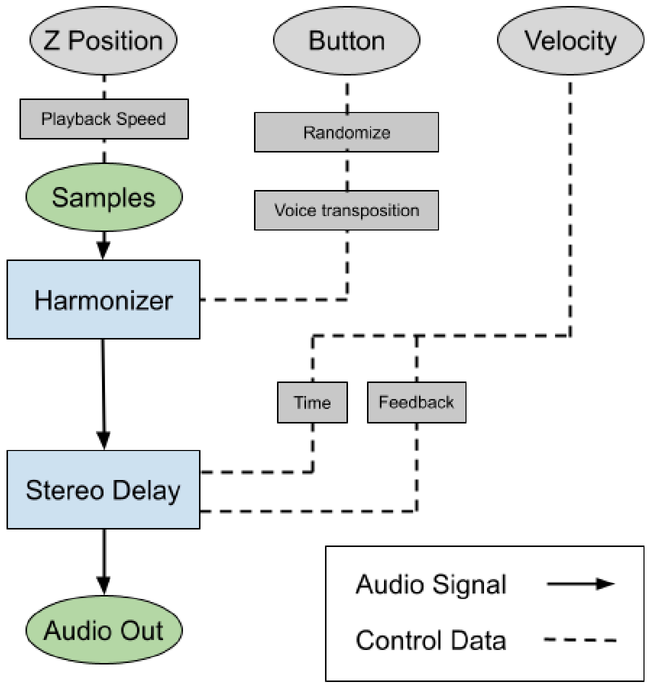


Figure 41. Mapping #3, harmonizer + stereo delay.

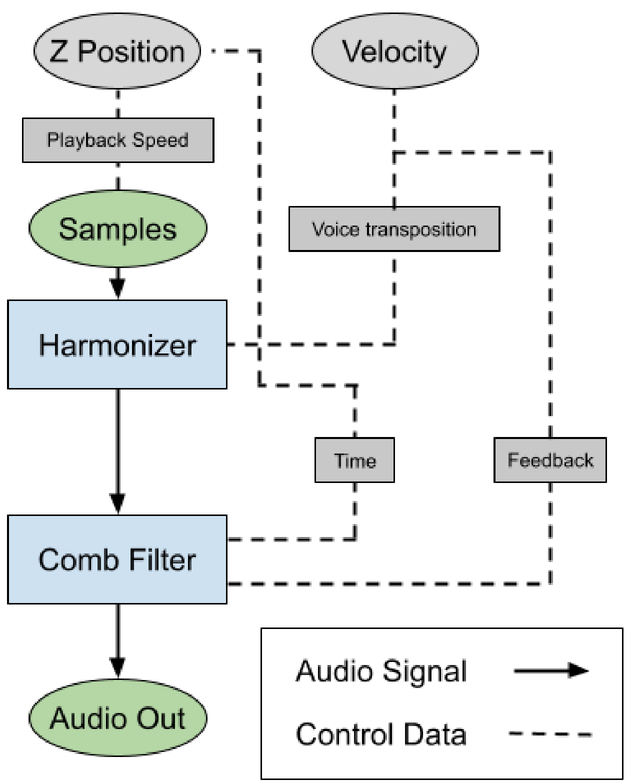


Figure 42. Mapping #4, harmonizer + comb filter.

Each of the four mappings has a distinct sound associated with it, with the first mapping being the one where the original sounds from the corpus are most recognizable. These mappings contribute in making the movement in virtual space more closely tied to the complex sonic outcome, allowing a performer to play the instrument beyond simply playing back the sound corpus.[360]

Chapter 3 Summary

In this chapter I have reviewed the different creative works that have been developed as part of my dissertation. Through these works I have been able to learn and explore the intersection of different tools for music production, computer music composition, spatial audio, 3D modeling, video game design, and interactive system design. The works trace a trajectory from the individual experience and embodied interaction towards interfacing with musicians performing musical instruments and the development of novel musical interfaces in virtual environments.

[360] The complete performance of *Point Cloud* and access to the 3D Corpus Navigator can be found in the appendix.

Conclusion

This document encapsulates an attempt to comprehend the possibilities and limitations in developing a musical practice in XR. I use ‘in’ XR and not ‘with’ XR because I have immersed myself into the history, theory, science, and creative works that have led musical XR to its current state. The research into these different aspects of music in XR have served as a lens to analyze, and contextualize the creative works that were developed as part of this dissertation.

The design of interactive musical systems using XR technology opens up multiple questions for the future development of computer music. Working in XR requires a variety of skills, including: 3D modeling, graphic design, and video game design. All of which are skills that go outside of what is usually taught in computer music, echoing Ciciliani’s idea of music in the expanded field.[361] In musical XR the relationship between time and space becomes more malleable, where virtual space can be constructed to dictate ‘musical’ time by determining the dimensions of virtual space and the spatial arrangements of interactive virtual objects in virtual and augmented spaces. Curtis Bahn uses an interactive adventure game for PC as an example of a meta-composition (Chapter 3). Bahn describes how every place in the game has a unique set of sounds that correspond with the location and its virtual inhabitants and how the actions of the player are linked to specific sounds as well. Bahn discusses its implications and potential for musical form, which can be found in the agency of the player, and the spatial/sonic couplings in the virtual environment.[362] The musical form in this example, or the possible musical forms that this game enables, are dictated by the arrangement of multiple virtual spaces within the video game, each with a unique soundscape, and unique actions available to the player. Bahn’s example is limited to the player, their actions, and their position in virtual space. Hamilton extends Bahn’s example by including virtual autonomous agents, third-party actors, and the inherent rules that govern the behavior

[361] Ciciliani, Marko. "Music in the expanded field: On recent approaches to interdisciplinary composition." *Darmstädter Beiträge zur Neuen Musik* 24 (2017): 23-35.
[362] Bahn, Curtis Robert. *Composition, improvisation and meta-composition*. Princeton University, 1997.

of virtual objects as mediation layers that play a role in generating music in virtual environments.[363] Virtual and augmented spaces contain the interactive system and all of its mediation layers and mapping schemes to generate sound, while it is the performer who interacts through the mediation layers in virtual space to articulate sounds and create music over time. In this context, virtual space can be used to control complex sound generating algorithms based on embodied interactive designs that exploit the spatial relationships between users and virtual objects, in co-located physical spaces, or across networked virtual environments. Music composition in XR then could be understood as the creation and organization of mediation layers that enable interactions from the performer to generate different musical outcomes.

Composers working in XR can organize multiple forms of media besides sound, including video, text, 3D animations, and images; they can establish any sort of relationships between these media through the design of interactive objects, virtual environments, autonomous agents, and the rules that govern the interactions between all these elements.[364] A better way to portray music composition in XR that includes the other forms of media besides sound, might be the analogy Lev Manovich uses in the *Language of New Media.* He sees the design of interactive computer media as fitting a trend of externalizing and objectifying the 'mind's operations', where users are "asked to follow pre-programmed, objectively existing associations. (...) Interactive media, ask us to identify with someone else's mental structure."[365] Music composition in XR reflects the mental structure of the composer in how they have organized and created layers of mediation between users, autonomous agents, virtual objects, images, animations, and the virtual environment to play or direct sound. For Manovich, the choice of a particular interface in new media art is motivated by a work's content, to the point where they cannot be thought of as separate entities. This idea resonates with the concepts reviewed in Chapter 2 of composed instrument and meta-composition. In the notion of a composed instrument, the musical instrument (interface) is also the musical score (content), and with meta-composition the media construct (interface) is what facilitates the

---

[363] Hamilton, Rob. "Mediated musical interactions in virtual environments." *New directions in music and human-computer interaction* (2019): 243-257.
[364] Hamilton (2019)
[365] Manovich, Lev. "The language of new media." (2002).

music performance/composition (content). Manovich's collapse of the interface and content in new media art is a useful lens to approach and understand music composition in XR, where what is being 'composed' is not only the music, but the experience to compose music in the relationships between the performer and the interactive system. To quote Rokeby in creating interactive art, "rather than creating finished works, the interactive artist creates relationships."[366] Something similar arises in music composition in XR.

The creation of virtual environments, its mediation layers, and the media it contains do not always need to drive the sound generating process. The organization of interactive virtual objects in virtual space can also be thought of as generating macro structures for real world musicians playing acoustic and/or electroacoustic instruments. The shapes, textures, colors, and movements in virtual space present opportunities to create 3D graphic scores, where virtual conductors can guide music performers through virtual space as they interpret the graphical elements musically.

Another major challenge in developing a musical practice in XR is the question of how to best present these works to audiences. This is both a problem that requires technical and artistic solutions in how to best integrate the virtual and real experiences of performers and audiences. The separation or blending together of virtual and real spaces, can be a fundamental part of the creative work, or simply a technical solution in order to create a better shared experience between performers and users. Some works might be intended to be experienced 'alone', having audiences be the performer and be immersed in the system, while others might seek both audiences and performers to go into a virtual space and interact through avatars.

Finally, it is important to consider how the integration of these technologies into our artistic practices tend to reflect back our vision of the world and our culture, for as George Lewis points out, the development of musical computer programs are not 'objective' or 'universal', similar to any 'texts' they represent particular ideas of their creators.[367] My hope is that moving forward with music and XR we can continue to integrate human qualities into these interactive systems through movement-based and

[366] Rokeby, David. "Transforming mirrors." *Leonardo Electronic Almanac* 3.4 (1995): 12.
[367] Lewis, George E. "Too many notes: Computers, complexity and culture in voyager." *Leonardo music journal* 10 (2000): 33-39.

embodied interactions, collaborative virtual environments, novel virtual musical instruments, audience participation, and mixed reality performances with real world musicians. The creative works presented in this dissertation are only a first step into reflecting and centering some of these ideas.

Appendix

Links to compositions, performances and demos

VR-Mapper demo 1: live input + granular delay
https://www.youtube.com/watch?v=zT9m1LSks1M

VR-Mapper demo 2: FM synthesizer
https://www.youtube.com/watch?v=Ruq2c8Xjtv8

UX-1
https://www.youtube.com/watch?v=LSZ7ALxAk4A

Push & Pull
https://www.youtube.com/watch?v=XL6pVQ2yRmE

pedalGround
https://www.youtube.com/watch?v=y6hexbgoZ7A

Unforeseen Collisions performed by Anıl Çamcı and Matias Vilaplana Stark
https://www.youtube.com/watch?v=F-EAICyC5ck

Senderos performed by New Thread Quartet
https://www.youtube.com/watch?v=GDNjfMKMZGs

Paisajes Oníricos / Dreamscapes performed by Popebama
https://www.youtube.com/watch?v=PWX_CnPkJYE

Paisajes Oníricos / Dreamscapes performed by Matias Vilaplana Stark
https://www.youtube.com/watch?v=OG3mhEU0n2w

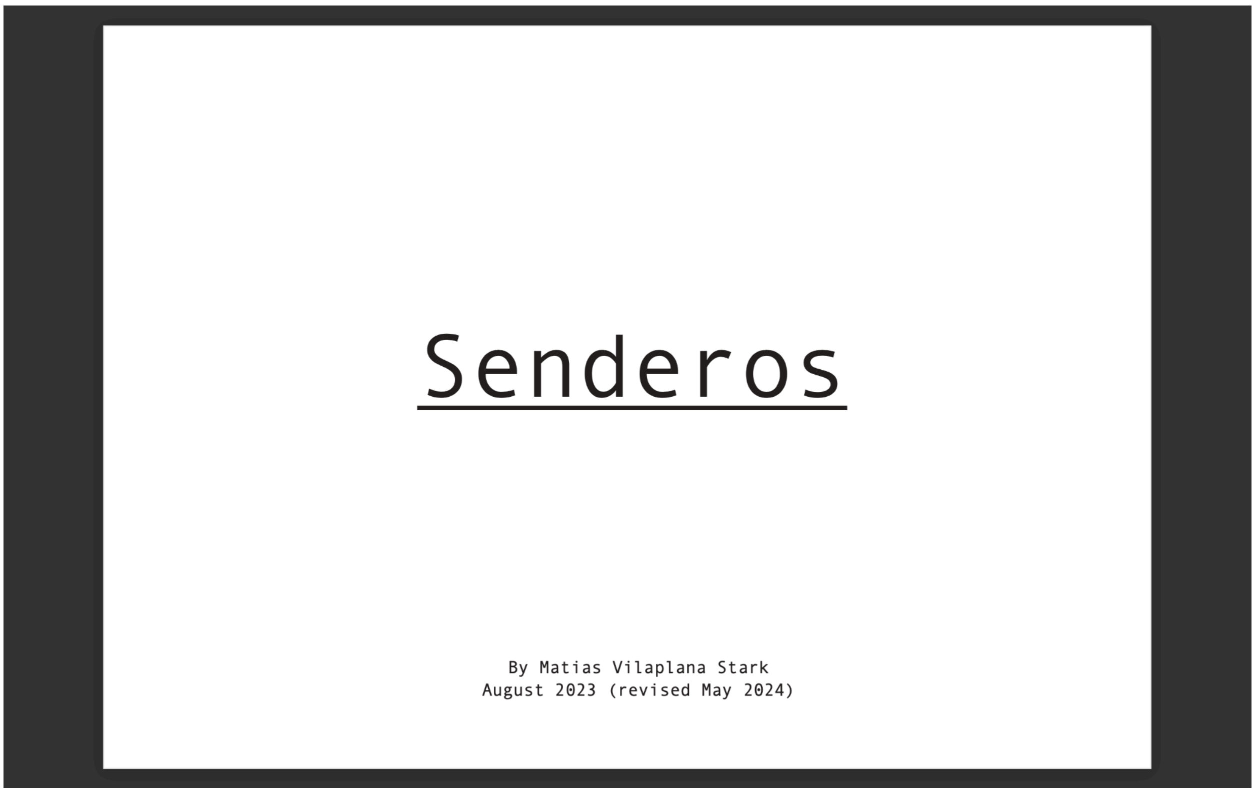
Funeral For A Whale (Stereo/binaural mix)
https://soundcloud.com/matiasvilaplana/funeral-for-a-whale

Point Cloud performed by Matias Vilaplana Stark
https://www.youtube.com/watch?v=gJVo3uGK-lQ

Scores

Senderos

[Link to video score](https://youtu.be/M589JBf7wYY)
https://youtu.be/M589JBf7wYY



Program Notes

I have not been able to travel to these places again, and now they only live in my memory. This piece is an attempt to remember through performance, to artificially travel through the digital representation of a place. Memory can never account for the complete experience and so digital representations can never be the real place.

Performance Notes

This version of Senderos has been arranged for a saxophone quartet, but it can be performed by any number and combination of instruments. It consists of a video score in which four digital landscapes are presented as different musical movements. In the following pages you will find sets of suggested music parameters to interpret each landscape.

Performers do not need to attain to all suggestions simultaneously, and should feel free to break, expand or go against the suggestions in favor of the musical outcome of the piece.

Colored bi-directional arrows are used to point out features of the landscape to look for in the video score:

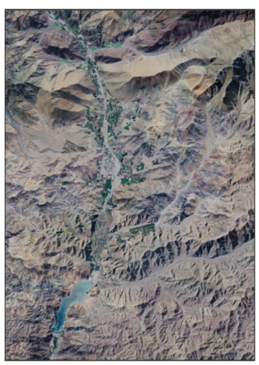
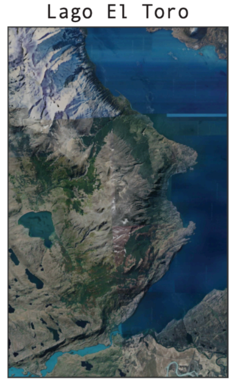
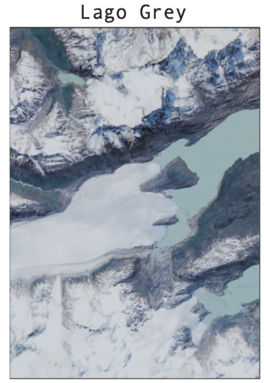
i.e. Water ← Terrain → Land

Black colored bi-directional arrows are used to represent musical parameters in a continuum and are associated with the terrain features of colored arrows:

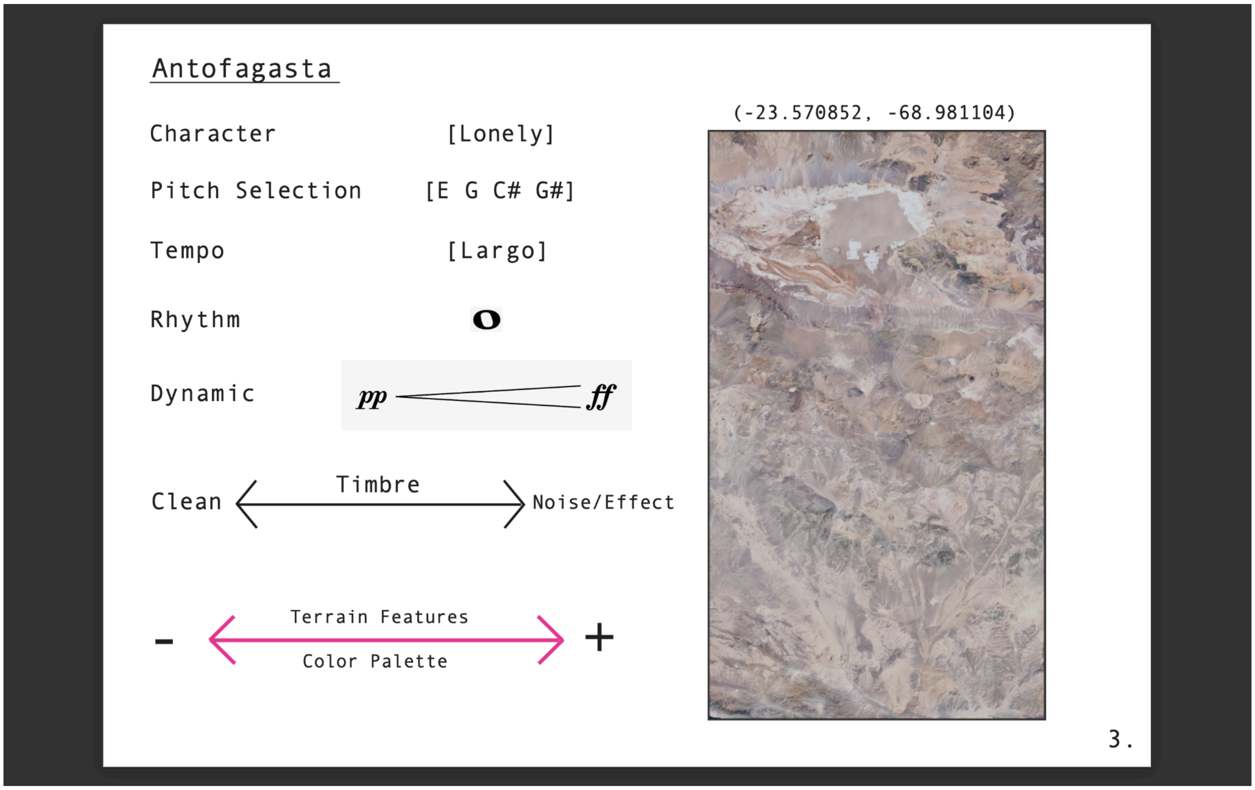
i.e. Rallentando ← Tempo → Accelerando

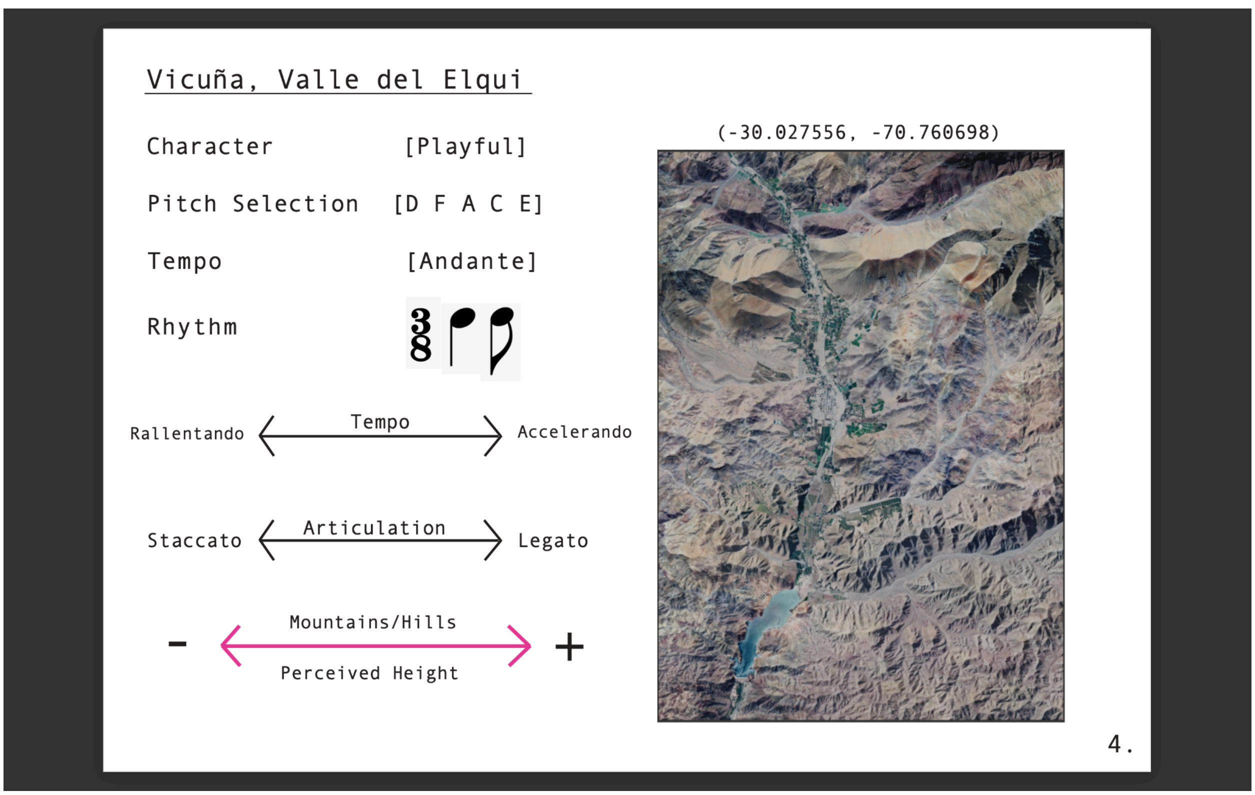
1.

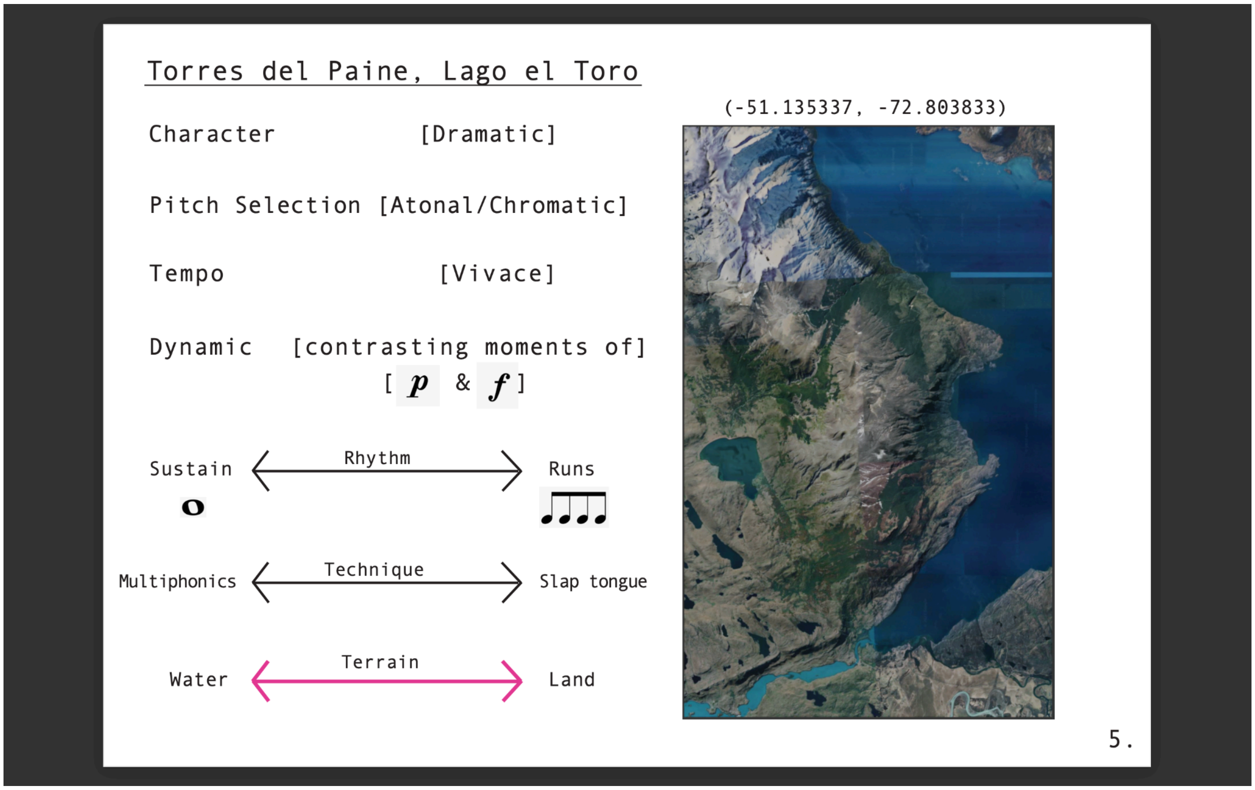
Video score time codes & landscapes

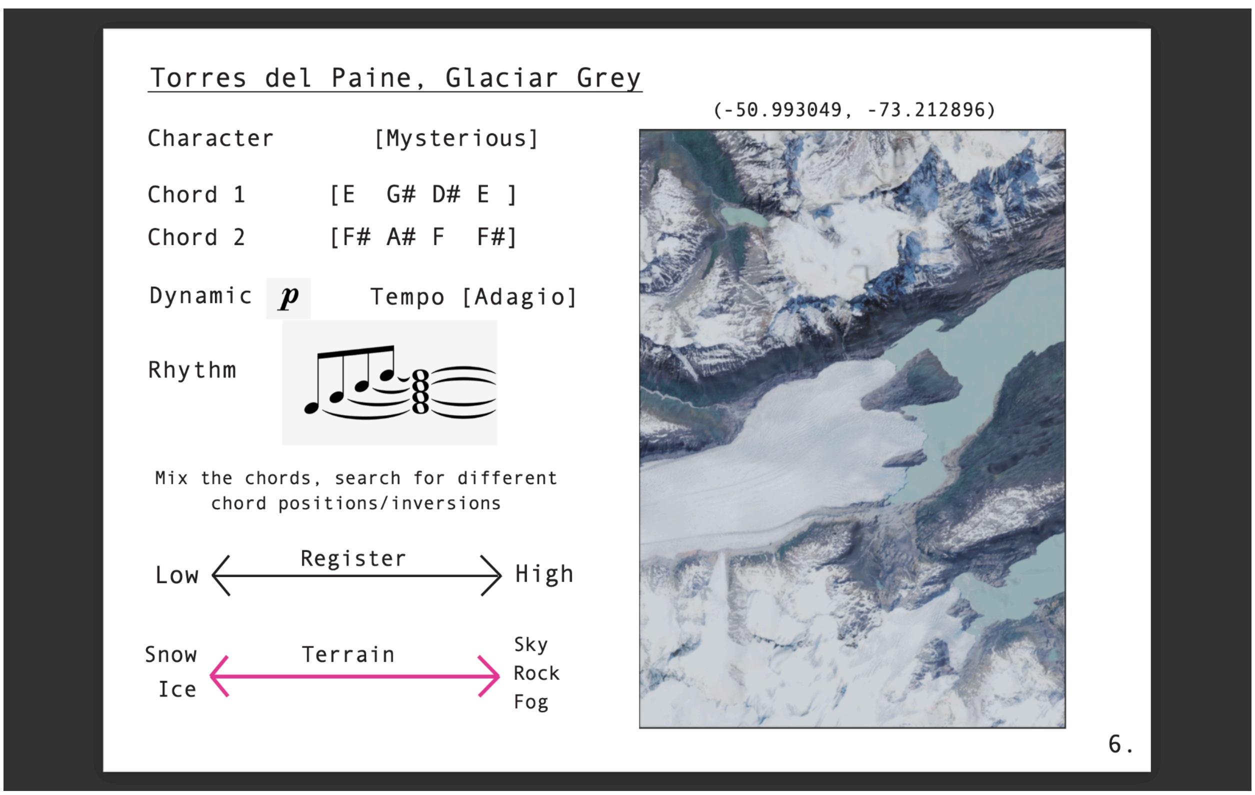
Antofagasta	Valle del Elqui	Torres del Paine, Lago El Toro	Torres del Paine, Lago Grey
			
00:00 - 02:10	02:10 - 05:25	05:25 - 07:50	07:50 - 10:20
Arid Cliffs No vegetation	Mountainous Semi-arid Agriculture	Windy Rain Forests	Cold Windy Ice Fields
(-23.570852, -68.981104)	(-30.027556, -70.760698)	(-51.135337, -72.803833)	(-50.993049, -73.212896)

2.


Antofagasta
Character [Lonely]
Pitch Selection [E G C# G#]
Tempo [Largo]
Rhythm
Dynamic
pp
ff
(-23.570852, -68.981104)
Clean
Timbre
Noise/Effect
-
Terrain Features
Color Palette
+
3.


Vicuña, Valle del Elqui
Character [Playful]
Pitch Selection [D F A C E]
Tempo [Andante]
Rhythm
3
8
(-30.027556, -70.760698)
Rallentando
Tempo
Accelerando
Staccato
Articulation
Legato
-
Mountains/Hills
Perceived Height
+
4.


Torres del Paine, Lago el Toro
(-51.135337, -72.803833)
Character [Dramatic]
Pitch Selection [Atonal/Chromatic]
Tempo [Vivace]
Dynamic [contrasting moments of]
[p & f]
Sustain
Rhythm
Runs
Multiphonics
Technique
Slap tongue
Water
Terrain
Land
5.


Torres del Paine, Glaciar Grey
(-50.993049, -73.212896)
Character [Mysterious]
Chord 1 [E G# D# E]
Chord 2 [F# A# F F#]
Dynamic p
Tempo [Adagio]
Rhythm
Mix the chords, search for different chord positions/inversions
Low
Register
High
Snow
Ice
Terrain
Sky
Rock
Fog
6.

Paisajes Oníricos / Dreamscapes

Paisajes Oníricos
/
Dreamscapes

Matias Vilaplana Stark
August 2023

Program Notes

Paisajes Oníricos/Dreamscapes reflects on the traditional representation of graphic scores as drawings on a two-dimensional plane. The piece expands on this notion by allocating the graphical elements in a 3D virtual environment. These elements acquire volume and mass, in addition to the incorporation of a spatio-temporal dimension that is controlled by a virtual conductor.

1.

Performance Notes

In this piece, the virtual environment is the graphic score. A virtual conductor navigates through virtual space, directing the performers' attention towards the different graphical elements of the score.

The 3D space is divided in three sections, each representing a musical movement. Instructions on how to navigate and interpret the graphical elements in each section are presented in the following pages.

Duration ~10 - 12 minutes

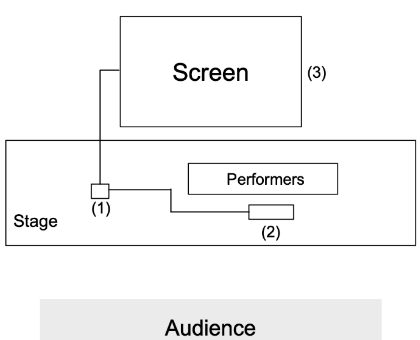
Performers/Instrumentation

- Virtual conductor
- Max loop patch, see page 14
- Open to any number of performers and any type of instruments: pitched, percussive* and electronics.

* Ideally an array of percussion instruments and/or objects of different materials (e.g. wood, metal, ceramic, skin, etc). Whichever the selection, it should aim to cover most of the hearing spectrum, from low to high.

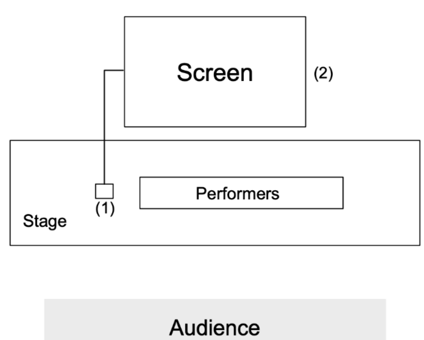
Stage Setup

Setup #1



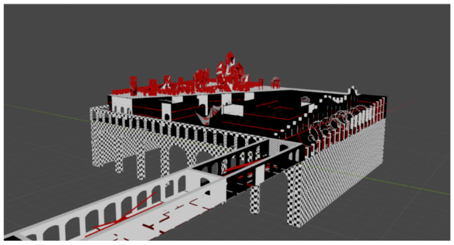
The virtual conductor computer (1) is connected to the performers' screen (2) and audience screen (3). Performers face in direction of the audience.

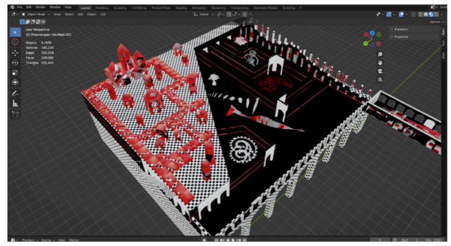
Setup #2

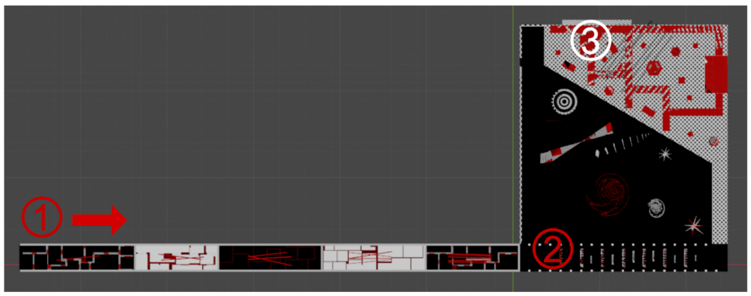


The virtual conductor computer (1) is connected to the audience screen (2) only. Performers face in direction of the screen, giving their back to the audience.

Virtual Environment







Top view/floor plan of the virtual world, composed of three sections.

4.

Virtual Conductor - Access & Controls

The virtual conductor navigates the virtual scene controlling the pacing, form and duration of the piece as they go through the different sections of the virtual environment.

You will need a computer with internet connection, a keyboard and a mouse (recommended over a trackpad). The controls are similar to those of a computer FPS game.

You can access the virtual environment using this link:

https://hubs.mozilla.com/scenes/cV4XgTH/paisajes-oniricos

Select "Create a room with this scene", and after it loads, select "Join Room". Choose your avatar, and you do not need to have either the microphone nor speakers on. You will be warped to the starting point of the virtual environment.

<u>Controls</u>

1) "WASD" or "Arrow" keys to move at walking speed, hold Shift to run:

    W - front

    A - left

    S - back

    D - right

2) Left click and drag on the screen to rotate your view.

3) Press the "B" key to enter "Record mode" which removes the Mozilla Hubs interface from the screen.

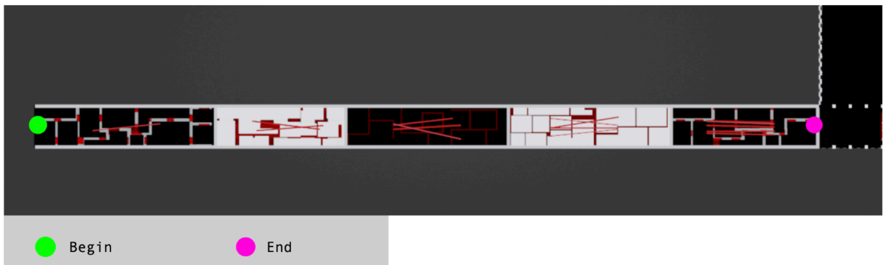
5.

Virtual Conductor - Instructions

Movement #1

The conductor is spawned looking at a white wall with the title of the piece. Turn your view 180° and walk through the long hall looking straight ahead. Avoid colliding into the red beams, walk around them if needed. This will allow performers to see the beams position change vertically as the conductor moves forward. This section ends when you reach a checkered gate.

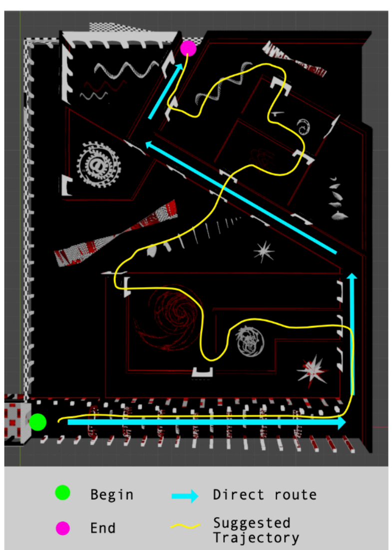
Top view



6.

Virtual Conductor - Instructions

Top view



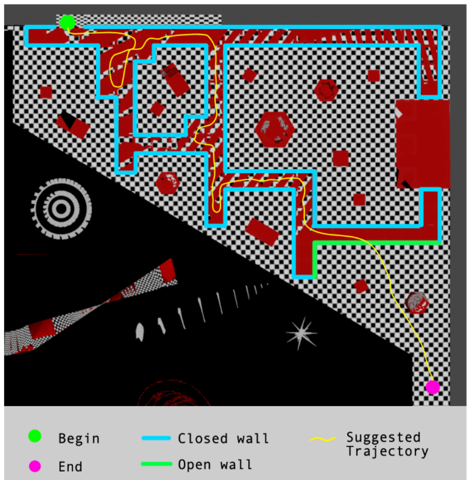
Movement #2

In this movement there are different paths the conductor can take. Throughout the scene the conductor will encounter a series of Vortexes (rotating structures of different forms and shapes). As you approach a Vortex, try to lock your view on it as you keep walking. This will allow performers to sonically interpret each Vortex until it is not possible to keep it in sight anymore. The end point for this movement is the checkered floor ramp that takes you to the second floor. A suggested trajectory is provided, but feel free to come up with a path of your own, or decide one in together with the ensemble.

7.

Virtual Conductor - Instructions

Top view



Movement #3

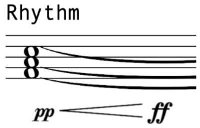
There should be silence as you move up the ramp to the second floor. The music should start the moment you enter the maze of red buildings. Hold shift and run through the maze, make the camera movement frantic, and try to keep it looking upwards towards the ceiling or the upper half of the scene rather than downwards. Do not stand still for too long. There are invisible walls that will prevent you from getting out of the maze. You can peek into the open space, but not enter it. The exit is at the SE part of the maze, you will see two buildings as you emerge into open space. As you walk away from the maze, the music should slowly become less frantic and fade out. Turn around and walk towards the end of the checkered floor while staring at the Maze, when you can not go any further (or when you think it is right), turn to your right to turn the screen black and end the piece.

8.

Movement#1 - Red Beams

Melodic motif

Rhythm

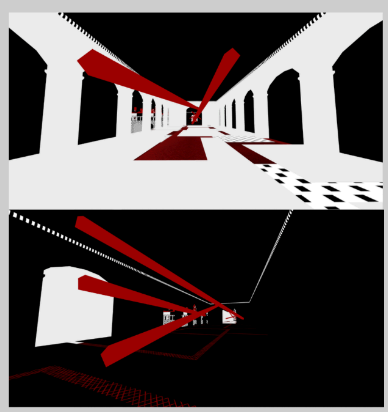


Take the melodic motif as a reference for selecting pitches to play in slow sustained tones.

Beam Vertical Position → Initial Pitch Register

Beam Vertical Trajectory → Glissando

Beam Length → Note Duration

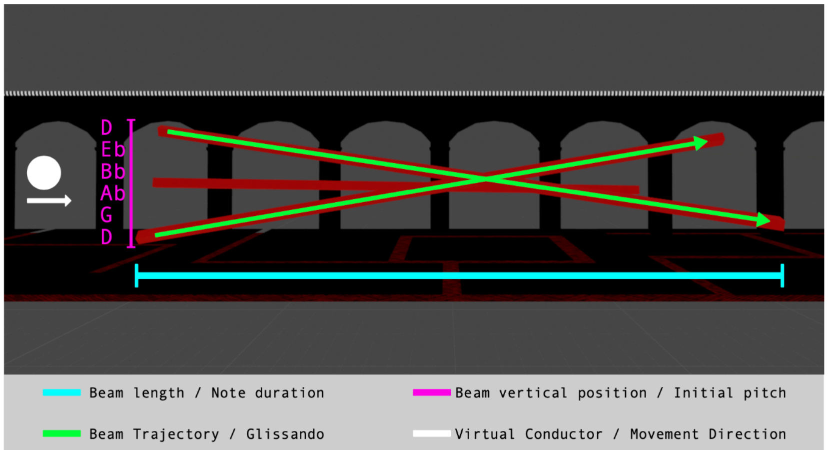


As the virtual conductor moves through the hall, long red beams slowly come into view, giving the impression of moving upwards or downwards.

Max Loop Patch, see page 14 for details

9.

Movement#1 - Red Beams



Even though a set of pitches are given in the melodic motif, these are not meant to be mapped exactly onto the vertical position of the beams. For example one could choose to interpret the vertical position of the lowest beam as a Bb instead of a D. You can decide this as an ensemble you or you can leave it open to individual interpretation. Similarly with the glissando, it is open to how much you think the trajectory of each beam represents (it could be a semi-tone, octave, etc).

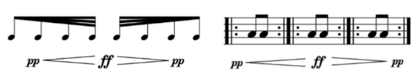
10.

Movement#2 - Vortex

Melodic motif

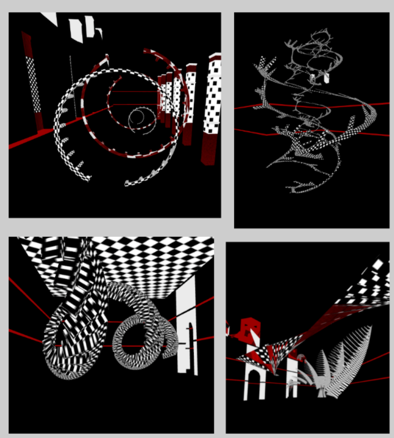
Mainly for pitch reference, but feel free to use this as you see fit.

Gestures (length is unspecified)



Vortex

- Rotation Speed → Gesture tempo
- Shape & Color → Type of articulation / Extended technique
- Proximity → Volume/Dynamic Range
- Volume & Mass → Register

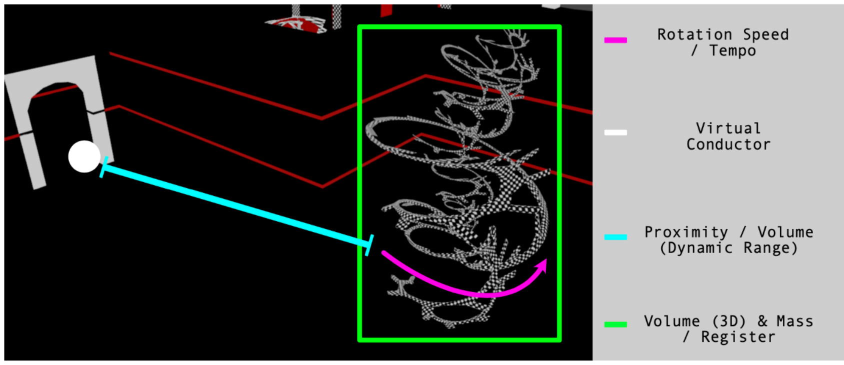


Structures of different shapes and color spin at different rates.

Try to coalesce your sounds as if they were all coming from the vortex in sight. This does not mean that your tempo needs to be synced, try to retain your own reading of tempo and explore poly-temporality as a group.

11.

Movement#2 - Vortex



In the previous page, it is suggested to associate volume and mass of the vortex to register. A vortex that is 'big' could suggest a sound on the lower register, but that same vortex, however big, might have thin branches that come out of it, which might be interpreted as a sound in the higher register. This associations can be open to individual interpretation or these can be decided collectively as you rehearse. A similar approach can be taken when it comes to deciding what types of articulation and extended techniques are more suitable to sonically interpret each vortex.

12.

Movement#3 - Maze

Chaotic, frenetic and loud, with short and abrupt moments of silence.

Follow the movement of the camera in portraying a sentiment of anxiety and frustration, as the conductor tries to escape the maze.

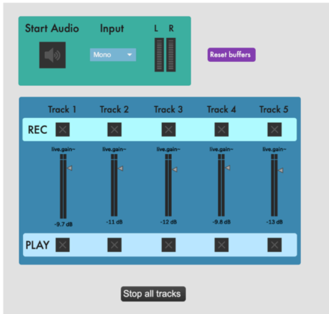
Let the shapes of black color in this environment suggest rhythm and pitch (atonal/noise). If the screen is covered in red, take that as a cue for silence.

Once the virtual conductor emerges from the maze, slowly decrease intensity and fade out until the screen is totally black.

The virtual conductor moves frenetically through a maze, trying to find their way out.

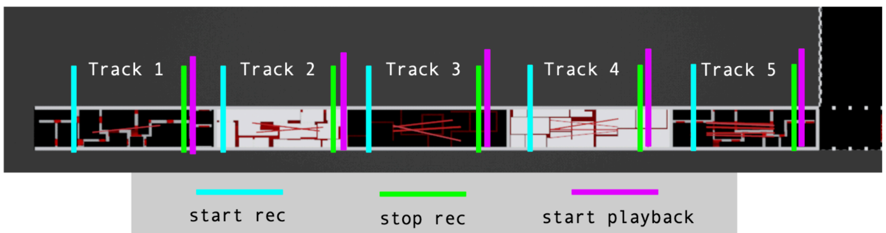
13.

Movement#1 – Max Loop Patch



The Max patch is a 5-track looping and playback device. Each track has a 25 second buffer and is used to record the 5 subsections of the first movement as shown in the figure below. Try to start recording before the performers produce sound and stop recording when they reach silence. Start the playback of each track right after you stop recording and slowly fade-in the track. The fade-in will probably happen while recording the next track.

The loops should keep playing throughout Movement #2 (balance mix accordingly) and be stopped right as Movement #3 starts.



14.

Bibliography

Abraham Gentile, Entertainment. “Power Glove.” (1989).

Agostinelli, Andrea, et al. “Musiclm: Generating music from text.” *arXiv preprint arXiv:2301.11325* (2023).

Alanko, Suvi. “Comparing Inside-out and Outside-in Tracking in Virtual Reality.” (2023).

Andersen, Michael Riis, et al. “Kinect depth sensor evaluation for computer vision applications.” *Aarhus University* (2012): 1-37.

Atherton, Jack, and Ge Wang. “Curating Perspectives: Incorporating Virtual Reality into Laptop Orchestra Performance.” *NIME.* 2020.

Atherton, Jack, and Ge Wang. “Doing vs. Being: A philosophy of design for artful VR.” *Journal of New Music Research* 49, no. 1 (2020): 35–59.

Bahn, Curtis Robert. *Composition, Improvisation and Meta-Composition.* Princeton University, 1997.

Barri, Tarik. “Versum: Audiovisual Composing in 3D.” (2009).

Beller, Greg. “Sound Space and Spatial Sampler.” *Proceedings of the 2nd International Workshop on Movement and Computing.* 2015.

Beller, Grégory. “Spatial Sampling in Mixed Reality.” *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA).* IEEE, 2023.

Beller, Grégory, and Georges Aperghis. “Gestural Control of Real-Time Concatenative Synthesis in Luna Park.” *P3S (Performative Speech and Singing Synthesis).* 2011.

Berger, Viktor. “Phenomenology of Online Spaces: Interpreting Late Modern Spatialities.” *Human Studies* 43.4 (2020): 603-626.

Berkhout, Augustinus J., Diemer de Vries, and Peter Vogel. “Acoustic control by wave field synthesis.” *The Journal of the Acoustical Society of America* 93.5 (1993): 2764-2778.

Berthaut, Florent. “3D Interaction Techniques for Musical Expression.” *Journal of New Music Research* 49.1 (2020): 60-72.

Berthaut, Florent. “Musical Exploration of Volumetric Textures in Mixed and Virtual Reality.” *Proceedings of the International Conference on New Interfaces for Musical Expression.* 2021.

Berthaut, Florent, et al. “First Person Shooters as Collaborative Multiprocess Instruments.” *New Interfaces for Musical Expression.* 2011.

Berthaut, F., Desainte-Catherine, M., and Hachet, M. “DRILE: an immersive environment for hierarchical live-looping.” *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)* (2010): 192–197.

Berthaut, Florent, Myriam Desainte-Catherine, and Martin Hachet. “Interacting with 3D Reactive Widgets for Musical Performance.” *Journal of New Music Research* 40.3 (2011): 253-263.

Berthaut, Florent, et al. “Reflets: Combining and Revealing Spaces for Musical Performances.” *New Interfaces for Musical Expression (NIME).* 2015.

Biocca, Frank. “The cyborg's dilemma: Progressive embodiment in virtual environments.” *Journal of Computer-Mediated Communication* 3.2 (1997): JCMC324.

Blanchard, Chuck, et al. “Reality built for two: a virtual reality tool.” *Proceedings of the 1990 symposium on Interactive 3D graphics* (1990), 35–36.

Bregler, Chris. “Motion capture technology for entertainment [in the spotlight].” *IEEE Signal Processing Magazine* 24.6 (2007): 160-158.

Botvinick, Matthew, and Jonathan Cohen. “Rubber hands ‘feel’ touch that eyes see.” *Nature* 391.6669 (1998): 756-756.

Boyer, Steven. “A virtual failure: Evaluating the success of Nintendo’s Virtual Boy.” *The Velvet Light Trap*, no. 64 (2009): 23–33.

Bowman, Doug A., and Ryan P. McMahan. “Virtual reality: how much immersion is enough?.” *Computer* 40.7 (2007): 36-43.

Butler, Jennifer. “Creating Pedagogical Etudes for Interactive Instruments.” *NIME.* 2008.

Cadoz, Claude, et al. “Responsive input devices and sound synthesis by stimulation of instrumental mechanisms: The cordis system.” *Computer Music Journal* 8.3 (1984): 60-73.

Camurri, Antonio. “On the role of artificial intelligence in music research.” *Journal of New Music Research* 19.2-3 (1990): 219-248.

Çamcı, Anıl, Matias Vilaplana, and Ruth Wang. “Exploring the Affordances of VR for Musical Interaction Design with VIMEs.” *Proceedings of the International Conference on New Interfaces for Musical Expression.* 2020.

Çamcı, Anıl, and John Granzow. “Hyperreal Instruments: Bridging VR and Digital Fabrication to Facilitate New Forms of Musical Expression.” *Leonardo Music Journal* 29 (2019): 14-18.

Çamcı, Anıl, and John Granzow. “Augmented Touch: A Mounting Adapter for Oculus Touch Controllers that Enables New Hyperreal Instruments.” (2022).

Caillon, Antoine, and Philippe Esling. “RAVE: A variational autoencoder for fast and high-quality neural audio synthesis.” *arXiv preprint* arXiv:2111.05011 (2021).

Chowning, John M. “The synthesis of complex audio spectra by means of frequency modulation.” *Journal of the Audio Engineering Society* 21.7 (1973): 526-534.

Ciciliani, Marko. “Music in the expanded field: On recent approaches to interdisciplinary composition.” *Darmstädter Beiträge zur Neuen Musik* 24 (2017): 23-35.

Ciciliani, Marko. “Virtual 3D environments as composition and performance spaces.” *Journal of New Music Research* 49.1 (2020): 104-113.

Civit, Miguel, et al. “A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends.” *Expert Systems with Applications* 209 (2022): 118190.

Cruz-Neira, Carolina et al. “The CAVE: audio visual experience automatic virtual environment.” *Communications of the ACM* 35, no. 6 (1992): 64–73.

Corness, Greg. “The musical experience through the lens of embodiment.” *Leonardo Music Journal* 18 (2008): 21-24.

Dahl, Luke, and Federico Visi. “Modosc: a library of real-time movement descriptors for marker-based motion capture.” *Proceedings of the 5th International Conference on Movement and Computing*. 2018.

Davies, Char. “OSMOSE: Notes on being in immersive virtual space.” (1998): 65-74.

Davies, Char, and John Harrison. “Osmose: towards broadening the aesthetics of virtual reality.” *ACM SIGGRAPH Computer Graphics* 30.4 (1996): 25-28.

Davis, Simon, Keith Nesbitt, and Eugene Nalivaiko. “A systematic review of cybersickness.” *Proceedings of the 2014 Conference on Interactive Entertainment* (2014).

Deacon, Thomas, and Mathieu Barthet. “Spatial Design Considerations for Interactive Audio in Virtual Reality.” *Sonic Interactions in Virtual Environments*. Cham: Springer International Publishing, 2022. 181-217.

Demers, Joanna. “Discursive Accents in Some Recent Digital Media Works.” (2013).

Diederick C. Niehorster, Li Li, and Markus Lappe. “The accuracy and precision of position and orientation tracking in the HTC Vive virtual reality system for scientific research.” *i-Perception* 8, no. 3 (2017): 2041669517708205.

Dziwis, Damian, and Henrik von Coler. “The entanglement: Volumetric music performances in a virtual metaverse environment.” *Journal of Network Music and Arts* 5.1 (2023): 3.

Dziwis, Damian, Henrik Von Coler, and Christoph Pörschmann. “Orchestra: A toolbox for live music performances in a web-based metaverse.” *Journal of the Audio Engineering Society* 71.11 (2023): 802-812.

Farrow, Robert, and Ioanna Iacovides. “Gaming and the Limits of Digital Embodiment.” *Philosophy & Technology* 27 (2014): 221-233.

Florens, Jean-Loup, et al. “ERGOS: Multi-degrees of freedom and versatile force-feedback panoply.” *EuroHaptics 2004.* (2004).

Furness III, Thomas A. “The super cockpit and its human factors challenges.” *Proceedings of the Human Factors Society Annual Meeting.* Vol. 30. No. 1. *Sage CA: Los Angeles, CA: SAGE Publications,* 1986.

Gabor, Dennis. “Theory of communication. Part 1: The analysis of information.” *Journal of the Institution of Electrical Engineers-part III: Radio and Communication Engineering* 93.26 (1946): 429-441.

Gerzon, Michael A. “Periphony: With-height sound reproduction.” *Journal of the Audio Engineering Society* 21.1 (1973): 2-10.

Gerzon, Michael A. “Surround-sound psychoacoustics.” *Wireless World* 80.1468 (1974): 483-486.

Gibson, James J. “Theories of Perception.” (1951).

Gleasure, Rob, and Joseph Feller. “A rift in the ground: Theorizing the evolution of anchor values in crowdfunding communities through the Oculus Rift case study.” *Journal of the Association for Information Systems* 17.10 (2016): 1.

Godøy, Rolf Inge, and Marc Leman, eds. *Musical Gestures: Sound, Movement, and Meaning.* Routledge, 2010.

Graham-Knight, Kimberlee, and George Tzanetakis. “Adaptive music technology using the Kinect.” *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 2015.

Gutierrez, Nicholaus. “The ballad of Morton Heilig: on VR's mythic past.” *JCMS: Journal of Cinema and Media Studies* 62.3 (2023): 86-106.

Halac, Fede Camara, and Shadrick Addy. “PathoSonic: Performing Sound In Virtual Reality Feature Space.” *NIME*. 2020.

Hackbarth, Benjamin, et al. “Composing morphology: Concatenative synthesis as an intuitive medium for prescribing sound in time.” *Contemporary Music Review* 32.1 (2013): 49-59.

Hamilton, Robert. “maps and legends: FPS-Based Interfaces for Composition and Immersive Performance.” *ICMC* (2007).

Hamilton, Robert. “Maps and legends: Designing fps-based interfaces for multi-user composition, improvisation and immersive performance.” *International Symposium on Computer Music Modeling and Retrieval*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

Hamilton, Robert. “Building interactive networked musical environments using q3osc.” *Audio Engineering Society Conference: 35th International Conference: Audio for Games.* Audio Engineering Society, 2009.

Hamilton, Rob. “Musical sonification of avatar physiologies, virtual flight and gesture.” *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers 10.* Springer, 2014. 518–532.

Hamilton, Rob. “Coretet: a 21st century virtual interface for musical expression.” *14th International Symposium on Computer Music Multidisciplinary Research.* Springer, 2019.

Hamilton, Rob. “Trois Machins de la Grâce Aimante: A virtual reality string quartet.” *Proceedings of the 2019 International Computer Music Conference, New York.* 2019.

Hamilton, Rob, and Chris Platz. “Gesture-based collaborative virtual reality performance in carillon.” *Proceedings of the 2016 International Computer Music Conference* (2016).

Hamilton, Rob. “Mediated musical interactions in virtual environments.” *New Directions in Music and Human-Computer Interaction* (2019): 243-257.

Heilig, Morton L. “Stereoscopic-television apparatus for individual use.” *U.S. Patent No. 2,955,156.* 4 Oct. (1960).

Hiller, Lejaren, and Pierre Ruiz. “Synthesizing musical sounds by solving the wave equation for vibrating objects: Part 1.” *Journal of the Audio Engineering Society* 19.6 (1971): 462-470.

Hunt, Andy, and Marcelo M. Wanderley. “Mapping performer parameters to synthesis engines.” *Organised sound* 7.2 (2002): 97-108.

Jacquemin, Christian, et al. “Plumage: Design d'une interface 3D pour le parcours d'echantillons sonores granularises.” *Proceedings of the 19th Conference on l'Interaction Homme-Machine*. 2007.

Jensenius, Alexander Refsum. “Action-sound: Developing methods and tools to study music-related body movement.” (2007).

Ji, Shulei, Jing Luo, and Xinyu Yang. “A comprehensive survey on deep music generation: Multi-level representations, algorithms, evaluations, and future directions.” *arXiv preprint* arXiv:2011.06801 (2020).

Karplus, Kevin, and Alex Strong. “Digital synthesis of plucked-string and drum timbres.” *Computer Music Journal* 7.2 (1983): 43-55.

Kilteni, Konstantina, Raphaela Groten, and Mel Slater. “The sense of embodiment in virtual reality.” *Presence: Teleoperators and Virtual Environments* 21.4 (2012): 373-387.

Kilteni, Konstantina, Ilias Bergstrom, and Mel Slater. “Drumming in immersive virtual reality: the body shapes the way we play.” *IEEE Transactions on Visualization and Computer Graphics* 19.4 (2013): 597-605.

Kim-Boyle, David. “The 3-D Score.” *TENOR 2017: International Conference on Technologies for Music Notation and Representation.* 2017.

Kim-Boyle, David. “3D Notations and the Immersive Score.” *Leonardo Music Journal* 29 (2019): 39-41.

King, Barbara J. *How animals grieve*. University of Chicago Press, 2013.

Knight-Hill, Andrew. “Theatres of sounds: The role of context in the presentation of electroacoustic music.” *Scene* 6.2 (2018): 165-175.

Krueger, Myron W. “Responsive environments.” *Proceedings of the June 13-16, 1977, National Computer Conference* (1977), 423–433.

Krueger, Myron W. “Artificial reality.” (1983).

Lanier, Jaron. *A Vintage Virtual Reality Interview.* Accessed on August 28th, 2023. http://www.jaronlanier.com/vrint.html.

Lanier, Jaron. *The Sound of One Hand. Whole Earth Review* 79 (1993): 30–34.

Lanier, Jaron. *Virtual Instrumentation.* Accessed on August 28th, 2024. http://www.jaronlanier.com/instruments.html.

Lansky, Paul. *The Importance of Being Digital.* TART Foundation and Gaudeamus Foundation, 2004.

LaViola Jr, Joseph J. "A discussion of cybersickness in virtual environments." *ACM SIGCHI Bulletin* 32.1 (2000): 47-56.

Lemmon, Eric C. "Telematic Music vs. Networked Music: Distinguishing Between Cybernetic Aspirations and Technological Music-Making." *Journal of Network Music and Arts* 1.1 (2019): 2.

Leonard, James, et al. "A virtual reality platform for musical creation: GENESIS-RT." *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers 10.* Springer International Publishing, 2014.

Lewis, George E. "Interacting with latter-day musical automata." *Contemporary Music Review* 18.3 (1999): 99-112.

Lewis, George E. "Too many notes: Computers, complexity and culture in Voyager." *Leonardo Music Journal* 10 (2000): 33-39.

Lewis, George E. "Improvised music after 1950: Afrological and Eurological perspectives." *Black Music Research Journal* (1996): 91-122.

Lewis, George E. "Living with Creative Machines: An Improvisor Reflects." *AfroGEEKS: Beyond the Digital Divide* (2007): 83-99.

Lewis, George E. "The virtual discourses of Pamela Z." *Journal of the Society for American Music* 1.1 (2007): 57-77.

Lewis, George E., interview on Voyager, *Cycling '74,* (2020). https://cycling74.com/articles/an-interview-with-george-lewis-and-damon-holzborn-part-1

Lewis, George E. "Co-creation: early steps and future prospects." In Bernard Lubat, Gérard Assayag, Marc Chemillier. *Artisticiel/Cyber-Improvisations.* Phonofaune, 2021, *Dialogiques d'Uzeste* (2021).

Li, Song, and Jürgen Peissig. "Measurement of head-related transfer functions: A review." *Applied Sciences* 10.14 (2020): 5014.

Loveridge, Ben. "An Overview of Immersive Virtual Reality Music Experiences in Online Platforms." *Journal of Network Music and Arts* 5.1 (2023): 5.

Lyon, Eric. "Spatial orchestration." *Proceedings of the 5th Sound and Music Computing Conference.* 2008.

Lyon, Eric. "The Future of Spatial Computer Music." *ICMC.* 2014.

MacKay, John. "On the perception of density and stratification in granular sonic textures: An exploratory study." *Journal of New Music Research* 13.4 (1984): 171-186.

Manovich, Lev. *The Language of New Media.* MIT Press, 2002.

Mann, Steve, et al. “All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality.” *arXiv preprint arXiv:1804.08386* (2018).

Manning, Peter. *Electronic and Computer Music.* Oxford University Press, USA (2013).

Marquez-Borbon, Adnan, and Juan Pablo Martinez Avila. “The problem of DMI adoption and longevity: Envisioning a NIME performance pedagogy.” (2018).

Masu, Raul, et al. “VR open scores: Scores as inspiration for VR scenarios.” *Proceedings of the International Conference on New Interfaces for Musical Expression, NIME 2020.* Birmingham City University, 2020. 109–114.

Mathews, Max V. “The Digital Computer as a Musical Instrument: A computer can be programmed to play 'instrumental' music, to aid the composer, or to compose unaided.” *Science* 142.3592 (1963): 553-557.

Mäki-Patola, Teemu, et al. “Experiments with virtual reality instruments.” *Proceedings of the 2005 conference on New Interfaces for Musical Expression.* 2005.

McCall, Lauren, and Jason Freeman. “A 3D Graphic Score Space and the Creative Techniques and Performance Practices that Emerge From It.” *Proceedings of the 16th International Audio Mostly Conference.* 2021.

McCauley, Michael E., and Thomas J. Sharkey. “Cybersickness: Perception of self-motion in virtual environments.” *Presence: Teleoperators & Virtual Environments* 1.3 (1992): 311-318.

Medeiros, Carolina Brum, and Marcelo M. Wanderley. “A comprehensive review of sensors and instrumentation methods in devices for musical expression.” *Sensors* 14.8 (2014): 13556-13591.

Mehri, Soroush, et al. “SampleRNN: An unconditional end-to-end neural audio generation model.” *arXiv preprint* arXiv:1612.07837 (2016).

Men, Liang, and Nick Bryan-Kinns. “LeMo: Supporting collaborative music making in virtual reality.” *2018 IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE).* IEEE, 2018.

Milgram, Paul, and Fumio Kishino. “A taxonomy of mixed reality visual displays.” *IEICE TRANSACTIONS on Information and Systems* 77.12 (1994): 1321-1329.

Minsky, Marvin. “Telepresence.” (1980).

Miranda, Eduardo Reck, and Marcelo M. Wanderley. *New digital musical instruments: control and interaction beyond the keyboard*. Vol. 21. AR Editions, Inc., 2006.

Mondloch, Katie. "A Symphony of Sensations in the Spectator: Le Corbusier's *Poeme Electronique* and the Historicization of New Media Arts." *Leonardo* 37.1 (2004): 57-62.

Mor, Noam, et al. “A universal music translation network.” *arXiv preprint* arXiv:1805.07848 (2018).

Møller, Henrik. “Fundamentals of binaural technology.” *Applied Acoustics* 36.3-4 (1992): 171-218.

Neupert, Max. *Exploring Concatenative Synthesis Units in VR.* Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2017.

Norman, Katharine. “Real-world music as composed listening.” *Contemporary Music Review* 15.1-2 (1996): 1-27.

Oliver, Julian, and Steven Pickles. “Fijuu2: A game-based audio-visual performance and composition engine.” *Proceedings of the 7th International Conference on New Interfaces for Musical Expression.* 2007.

Ortega, Basilio Pueo, and José Manuel Jiménez Olmedo. “Application of motion capture technology for sport performance analysis.” *Retos: Nuevas tendencias en educación física, deporte y recreación* 32 (2017): 241-247.

Özcan, Zeynep, and Anil Çamci. “An augmented reality music composition based on the sonification of animal behavior.” *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society, 2018.

J. T. Reason and J. J. Brand. *Motion Sickness.* London: Academic Press, 1975.

Palumbo, Michael, Alexander Zonta, and Graham Wakefield. “Modular reality: Analogues of patching in immersive space.” *Journal of New Music Research* 49.1 (2020): 8-23.

Paradiso, Joseph A., and James A. Landay. “Guest editors' introduction: Cross-reality environments.” *IEEE Pervasive Computing* 8.3 (2009): 14-15.

Parth Rajesh Desai, et al. “A review paper on Oculus Rift—a virtual reality headset.” *arXiv preprint arXiv:1408.1173* (2014).

Păvăloiu, Ionel-Bujorel. “Leap motion technology in learning.” *Edu World 7th International Conference.* 2017.

Peddie, Jon. “Historical overview: ghosts to real AR to DARPA.” *Augmented Reality: Where We Will All Live.* Cham: Springer International Publishing, (2023): 101-133.

Perkis, Tim, et al. *The League of Automatic Music Composers, 1978-1983.* New World Records, 2007.

Politis, Archontis. “Microphone array processing for parametric spatial audio techniques.” (2016).

Pulkki, Ville. “Virtual sound source positioning using vector base amplitude panning.” *Journal of the Audio Engineering Society* 45.6 (1997): 456-466.

Renganayagalu, Sathiya Kumar, Steven C. Mallam, and Salman Nazir. “Effectiveness of VR head-mounted displays in professional training: A systematic review.” *Technology, Knowledge and Learning* (2021): 1-43.

Risset, Jean-Claude. “Computer study of trumpet tones.” *The Journal of the Acoustical Society of America* 38.5_Supplement (1965): 912-912.

Risset, Jean-Claude. “Real-world sounds and simulacra in my computer music.” *Contemporary Music Review* 15.1-2 (1996): 29-47.

Risset, Jean-Claude. “Computer music: why.” *Internet Proceedings of Composers' Forum Austin* (2003).

Roads, Curtis. “Artificial intelligence and music.” *Computer Music Journal* 4.2 (1980): 13-25.

Roads, Curtis. “Improvisation with George Lewis.” *Composers and the Computer* (1985): 75-88.

Roads, Curtis. “Research in music and artificial intelligence.” *ACM Computing Surveys (CSUR)* 17.2 (1985): 163-190.

Roads, Curtis. “Introduction to granular synthesis.” *Computer Music Journal* 12.2 (1988): 11-13.

Roads, Curtis. *Composing Electronic Music: A New Aesthetic.* Oxford University Press, 2015.

Rokeby, David. “Transforming mirrors.” *Leonardo Electronic Almanac* 3.4 (1995): 12.

Rottondi, Cristina, et al. “An Overview on Networked Music Performance Technologies.” *IEEE Access* 4 (2016): 8823-8843.

Sá, Adriana. “A Method for the Analysis of Sound Art and Audio-Visual Performance.” *Audiovisual e Industrias Criativas: Presente e Futuro* (2021): 575-589.

Sanchez-Vives, Maria V., et al. “Virtual hand illusion induced by visuomotor correlations.” *PloS One* 5.4 (2010): e10381.

Serafin, Stefania, et al. “Virtual reality musical instruments: State of the art, design principles, and future directions.” *Computer Music Journal* 40.3 (2016): 22–40.

Simon, Herbert A., and Richard K. Sumner. “Pattern in music.” *Formal Representation of Human Judgment.* New York: Wiley (1968): 219-250.

Schnell, Norbert, and Marc Battier. “Introducing composed instruments, technical and musicological implications.” *Proceedings of the 2002 conference on New interfaces for musical expression*. 2002.

Schwarz, Diemo, et al. “Real-time corpus-based concatenative synthesis with catart.” *9th International Conference on Digital Audio Effects (DAFx)*. 2006.

Schwarz, Diemo, et al. “Musical applications of real-time corpus-based concatenative synthesis.” *International Computer Music Conference (ICMC)*. 2007.

Slater, Mel. “Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009): 3549-3557.

Slater, Mel, et al. “Immersion, presence and performance in virtual environments: An experiment with tri-dimensional chess.” *Proceedings of the ACM Symposium on Virtual Reality Software and Technology.* 1996.

Slater, Mel, Bernhard Spanlang, and David Corominas. “Simulating virtual environments within virtual environments as the basis for a psychophysics of presence.” *ACM Transactions on Graphics (TOG)* 29.4 (2010): 1-9.

Slater, Mel, and Sylvia Wilbur. “A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments.” *Presence: Teleoperators & Virtual Environments* 6, no. 6 (1997): 603–616.

Smalley, Denis. “Spectromorphology: Explaining Sound-Shapes.” *Organised Sound* 2.2 (1997): 107-126.

Smith, J. Howard, and K. JH. “The sound field microphone.” *db Magazine* 12.7 (1978).

Spanlang, Bernhard, et al. “How to build an embodiment lab: achieving body representation illusions in virtual reality.” *Frontiers in Robotics and AI* 1 (2014): 9.

Stanney, Kay M., et al. “Extended reality (XR) environments.” *Handbook of Human Factors and Ergonomics* (2021): 782-815.

Steinbeck, Paul. “George Lewis’s Voyager.” *The Routledge Companion to Jazz Studies.* Routledge, 2018. 261-270.

Steinicke, Frank. *Being Really Virtual: Immersive Natives and the Future of Virtual Reality.* Springer, 2016.

Stine, Eli. “Creating Immersive Electronic Music from the Sonic Activity of Environmental Soundscapes.” *Iui workshops*. 2019.

Sutherland, Ivan E. “A head-mounted three dimensional display.” *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I* (1968), 757–764.

Sutherland, Ivan E. “The ultimate display.” *Proceedings of the IFIP Congress*. Vol. 2. No. 506-508. 1965.

Svanæs, Dag. “Interaction design for and with the lived body: Some implications of Merleau-Ponty's phenomenology.” *ACM transactions on computer-human interaction (TOCHI)* 20.1 (2013): 1-30.

Tremblay, Pierre Alexandre, Gerard Roma, and Owen Green. “Enabling Programmatic Data Mining as Musicking: The Fluid Corpus Manipulation Toolkit.” *Computer Music Journal* 45.2 (2021): 9-23.

Truax, Barry. “Real-time granular synthesis with a digital signal processor.” *Computer Music Journal* 12.2 (1988): 14-26.

Truax, Barry. “Genres and techniques of soundscape composition as developed at Simon Fraser University.” *Organised sound* 7.1 (2002): 5-14.

Turowski, Paul. "SQ2." (2019). https://livrepository.liverpool.ac.uk/3104992/

Van Den Oord, Aaron, et al. “Wavenet: A generative model for raw audio.” *arXiv preprint* arXiv:1609.03499 12 (2016).

Varèse, Edgard, and Chou Wen-Chung. “The Liberation of Sound.” *Perspectives of New Music* 5.1 (1966): 11-19.

Vercoe, Barry. “The synthetic performer in the context of live performance.” *Proceedings of International Computer Music Conference.* 1984.

Vilaplana Stark, M. A. “Developing a 3D interface for sound corpus manipulation in virtual environments,” *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Saint Malo, France, 2025, pp. 643-646.

Visi, Federico, Rodrigo Schramm, and Eduardo Miranda. "Gesture in performance with traditional musical instruments and electronics: Use of embodied music cognition and multimodal motion capture to design gestural mapping strategies." *Proceedings of the 2014 International Workshop on Movement and Computing*. 2014.

Vorländer, Michael. "Past, present and future of dummy heads." *Proc. Acústica, Guimarães, Portugal* (2004): 13-17.

Vredeman de Vries, Hans. *Perspective*. Dover Publications, 1968.

Wallraff, Dean. "The DMX-1000 signal processing computer." *Computer Music Journal* (1979): 44-49.

Webster, Rustin and Alex Clark. "Turn-key solutions: Virtual reality." In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 57052 (*American Society of Mechanical Engineers*, 2015), V01BT02A052.

Williams, Dmitri. "Structure and competition in the US home video game industry." *International Journal on Media Management* 4.1 (2002): 41-54.

Winograd, Terry. "Linguistics and the computer analysis of tonal harmony." *Journal of Music Theory* 12.1 (1968): 2-49.

Witmer, Bob G., and Michael J. Singer. "Measuring presence in virtual environments: A presence questionnaire." *Presence* 7.3 (1998): 225-240.

Wright, Matthew. "Open Sound Control: an enabling technology for musical networking." *Organised Sound* 10.3 (2005): 193-200.

Wozniak, Peter, et al. "Possible applications of the LEAP motion controller for more interactive simulated experiments in augmented or virtual reality." *Optics Education and Outreach IV.* Vol. 9946. SPIE, 2016.

Xenakis, Iannis. *Formalized Music: Thought and Mathematics in Composition.* No. 6. Pendragon Press, (1992).

Xiuquan Qiao, et al. "Web AR: A promising future for mobile augmented reality—State of the art, challenges, and insights." *Proceedings of the IEEE* 107, no. 4 (2019): 651–666.

Young-Geun Kim, and Won-Jung Kim. "Implementation of augmented reality system for smartphone advertisements." *International Journal of Multimedia and Ubiquitous Engineering* 9, no. 2 (2014): 385–392.

Zachara, Matt, and José P. Zagal. "Challenges for success in stereo gaming: a Virtual Boy case study." *Proceedings of the International Conference on Advances in Computer Entertainment Technology.* 2009.

Zappi, Victor, et al. "Design and Evaluation of a Hybrid Reality Performance." *NIME*. Vol. 11. 2011.

Zappi, Victor, et al. "Concatenative Synthesis Unit Navigation and Dynamic Rearrangement in VRGrains." *Proceedings of the Sound & Music Computing Conference.* 2012.

Zappi, Victor, Dario Mazzanti, and Florent Berthaut. *From the Lab to the Stage: Practical Considerations on Designing Performances with Immersive Virtual Musical Instruments.* 2022.

Zbyszynski, Michael. “An Elementary Method for Tablet.” *NIME*. 2008

Zbyszyński, Michael, et al. “Gesture-timbre space: Multidimensional feature mapping using machine learning and concatenative synthesis.” *International Symposium on Computer Music Multidisciplinary Research*. Cham: Springer International Publishing, 2019.

Zerpa, Carlos et al. “The use of Microsoft Kinect for human movement analysis.” *International Journal of Sports Science* 5.4 (2015): 120-127.

Zhu, Andrew, and Ge Wang. “ChuGL: Unified Audiovisual Programming in ChucK.” *Proceedings of the International Conference on New Interfaces for Musical Expression.* 2024.

Zotter, Franz, and Matthias Frank. *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality.* Springer Nature, 2019.