The Design and Analysis of Mixed Reality Musical Instruments

by

Karitta Christina (Kit) Zellerbach

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Interactive Media and Game Development

August 2022

APPROVED:

Professor Charles Roberts, Thesis Advisor

Professor Gillian Smith, Thesis Reader

Abstract

In the context of immersive sonic interaction, Virtual Reality Musical Instruments have had the relative majority of attention thus far, fueled by the increasing availability of affordable technology. Recent advances in Mixed Reality (MR) experiences have provided the means for a new wave of research that goes beyond Virtual Reality. In this paper, we propose a new classification of Virtual Musical Instrument, known as a Mixed Reality Musical Instrument (MRMI). We define this system as an embodied interface for expressive musical performance, characterized by the relationships between the performer, the virtual, and the physical environment. We offer a dimensional framework to support the analysis and design of MRMIs, illustrate its use with application to existing works, and evaluate it through expert interviews. These interviews highlight both the importance of the framework in a nascent domain and suggest further consideration of musical practice. Finally, we present *Wavelength*, a novel MRMI based on the metaphor of wave terrain synthesis, which we evaluate using audience perception of an improvised performance, personal reflection, and our proposed framework. We are able to conclude with a demonstrable appeal for musical performance in MR—a promising future for MRMIs.

Acknowledgements

I want to formally express my gratitude to my advisor, Charlie Roberts, for his continuous guidance and support throughout my studies at WPI. This thesis would certainly not exist without him, nor would I have been empowered to believe that I could succeed and make meaningful contributions to the field. I also want to thank my reader, Gillian Smith, not only for taking the time to provide feedback on my thesis, but also for being one of the most inspiring and caring people I had have the pleasure to meet. Thank you to all the expert interviewees and reviewers at NIME for their invaluable feedback. Thank you to Microsoft and the employee tuition reimbursement program that made this all the more possible. I remain forever grateful to my friends and family for their endless love, encouragement, and understanding when I had to cancel plans to code.

Contents

List of Figures

Abbreviations

VR Virtual Reality

AR Augmented Reality

MR Mixed Reality

XR Extended Reality

RV Reality-Virtuality (Continuum)

DMI Digital Musical Instrument

VMI Virtual Musical Instrument

VRMI Virtual Reality Musical Instrument

MRMI Mixed Reality Musical Instrument

HMD Head-mounted Display

HHD Hand-held Display

OST Optical See-Through

VST Video See-Through

DoF Degrees of Freedom

HCI Human Computer Interaction

WT Wave Terrain

AM Amplitude Modulation

FM Frequency Modulation

1 Introduction

In 1992, Jaron Lanier debuted *The Sound of One Hand*, a live musical performance in virtual reality [1]. Equipped with a Head-mounted Display (HMD) and single DataGlove, he played a selection of virtual musical instruments using only one hand. A feat made possible by virtualization, and one that paved the way for a new field of research in immersive sonic interaction.

While past researchers in the field of New Interfaces for Musical Expression (NIME) have focused on Virtual Reality Musical Instruments (VRMIs), interfaces for Mixed Reality (MR) have yet to receive the same attention. VRMIs almost necessarily involve cybersickness, a condition with a 40–60% susceptibility rate that disproportionately affects females [2][3], explained further in section 4.2.1. While the effects of cybersickness can be mitigated through careful design practice, such considerations often place significant restrictions on movement, limiting the potential of the VR medium. High-end equipment with lower latencies, mechanical optical adjustments, and a wider field of view can also help with cybersickness, although the cost of such systems is comparable to current HMD-based MR technologies and arguably out out of reach for the typical consumer.

Our contribution to this field is two-fold. First, we explore existing VRMI evaluation frameworks and design principles. We consider the affordances provided in MR systems and compare these to their VR counterparts. We define a new wave of virtual instrumentation as Mixed Reality Musical Instruments (MRMIs), which differ in fundamental ways from their VRMI predecessors. Finally, we offer a framework that enables NIME practitioners to design and analyze MRMIs through a common shared language. Rather than replace these existing frameworks, we seek to both refine and extend them into the realm of mixed user-reality-virtuality interaction. We apply this framework to analyze existing musical performances and applications in MR. After being peer-reviewed, this work was published in the proceedings of the 2022 New Interfaces for Musical Expression conference [4]. We detail some of the changes made based on reviewer feedback and evaluate the framework through discussion from expert interviews.

Second, we develop *Wavelength*, an instrument for expressive audiovisual performance with the Microsoft HoloLens. The system consists of malleable holographic objects, as polymorphic terrains, which performers will reshape and orchestrate to create generative music. We evaluate this work through audience perception of an improvised performance, as well personal reflection as both performer and designer. Finally, we offer discussion through the dimensions of our proposed framework.

2 Extended Reality

In this section, we explore paradigms of Extended Reality (XR) systems, an umbrella term for AR, MR, and VR systems. We offer definitions and outline the prominent tools and technologies involved.

2.1 Virtual Reality (VR)

Virtual reality (VR) is a simulated experience where the participant is immersed in a synthetic world that can be similar to or different from the real world [5].

In 1992, Cruz *et al.*, noted that modern virtual reality research had split into four distinct directions, based primarily on differences in display devices [6]. These are Cathode Ray Tube (CRT), Head-Mounted Display (HMD), Binocular Omni-Oriented Monitor (BOOM), and Cave Automatic Virtual Environment (CAVE).

2.2 Augmented Reality (AR)

Augmented Reality (AR) at its core refers to the experience of a real-world environment that is enhanced by computer-generated perceptual information. This information can span across multiple sensory modalities, such as visual, auditory, and haptic. AR systems are defined by their incorporation of three key characteristics [7]:

- the combination of the real and virtual
- real-time interactivity
- 3D registration

2.2.1 Hand-held Display (HHD)

More commonly, mobile devices are used as AR displays, where users utilize their screen as a lens into the real word, with augmentations based on camera and sensor input [8][9]. Their accessibility and ubiquity provides the experience of AR to the masses. With more common applications in social media, users of the popular messaging app Snapchat can share photos with visual augmentations, as shown in Figure 1.

Figure 1: A *Snapchat Lens* filter utilizing AR technology

Hand-held AR technology makes use of monitor-based AR displays, a non-immersive experience in which computer graphic (CG) images are overlaid on a screen, such as a TV or smartphone. The display is treated as a window to the augmented world, hence the alternative name "Window-on-the-World" [5].

2.2.2 Head-mounted Display (HMD)

Head-mounted displays are most commonly associated with AR and VR experiences. Early AR exploration was formed around the use of HMDs for aircraft manufacturing at Boeing. The HMD as a model of AR still exists in the nostalgia of "yesterday's tomorrows" [10]. There are two types of see-through (ST) HMD-based AR systems: optical and video.

Optical See-Through (OST) HMD With optical-see-through HMDs, the real world is seen through semi-transparent mirrors placed in front of the user's eyes. These mirrors are also used to reflect the computer generated images into the user's eyes, thereby combining the real- and virtual-world views. [11]. Examples of OST HMD devices include the Google Glass[1] and Microsoft HoloLens[2].

[1] `https://www.google.com/glass`
[2] `https://www.microsoft.com/en-us/hololens`

Video See-Through (VST) HMD With a video see-through HMD, the real-world view is captured with two miniature video cameras mounted on the head gear, and the computer-generated images are electronically combined with the video representation of the real world [11]. An example of this is the ZED Mini[3].

2.3 Mixed Reality (MR)

Mixed Reality (MR), as defined by Milgram and Kishino, exists as the spectrum between the two extremes of an entirely physical world and entirely virtual world [5]. They believe it is best illustrated by the Reality-Virtuality (RV) Continuum (see Figure 2), which defines a Mixed Reality environment as one in which real-world and virtual world objects are presented together within a single display, that is, anywhere between, but not on, the extrema of the virtuality continuum.

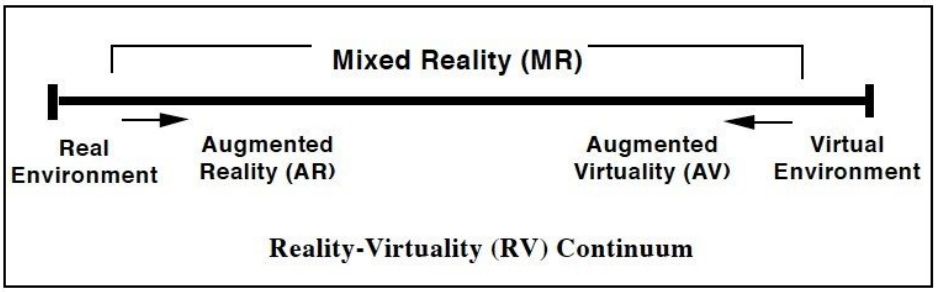


Figure 2: The Reality-Virtuality Continuum [5]

Milgram and Kishino offer a three-dimensional taxonomy for mixing real and virtual worlds:

- **Extent of World Knowledge (EWK)** describes how much we know about items and the environment in which they are shown.
- **Reproduction Fidelity (RF)** refers to the relative quality with which the synthesising display can recreate the real or intended pictures of the objects being presented.
- **Extent of Presence Metaphor (EPM)** describes the extent to which the observer is intended to feel "present" within the displayed scene.

According to the framework, VR is not part of MR while AR is considered a subclass of MR. While the difference between AR and VR is typically well-understood, the difference between AR and MR is relatively ambiguous as the terms are often used interchangeably. Interviews with domain experts [12] concluded in contradicting notions on what constitutes MR, with statements such as

[3] `https://www.stereolabs.com/zed-mini`

"the same as AR", "the RV continuum", and "technology-bound". One perception of MR is as an evolution of AR, distinguished by an advanced spatial understanding and interaction between users and virtual objects, and virtual objects with the environment. However, this idea may lead to the conclusion that MR is constrained to the hardware that is able to deliver this functionality. As such, associating definitions of XR systems with specific technologies can be problematic, though a practice that often appears in literature. The RV Continuum itself acts as a foundation for most AR/MR research, yet it is based on the idea that all MR experiences are best represented by their method of display. Indeed, the vast majority of AR research employs the use of an HMD. Many researchers have since moved on from this display-based taxonomy, expanding its definition to include other modalities of sensing, such as audio, proprioceptive, haptics, taste/flavor, and smell [12][13].

To this end, it may be useful to examine designations of MR that seek to remove technology as the foundation of their taxonomies. [14] offers a new class of MR systems, denoted as MR^x to mark the importance of user experience. MR^x applications are designed to be engaging experiences, rather than instruments for completing tasks. They are distinguished by three main qualities:

- hybrid
- deeply locative and often site-specific
- esthetic, performative, and/or social

First, MR^x applications are hybrid in the sense that they seek to combine the physical and virtual effectively for the sake of the experience, whether it is seamless integration or radical separation. Second, all MR applications are in some way location-based or location-aware. The difference is that for MR^x, relation to location is fundamental to the experience, either esthetically or culturally. In this sense, these experiences can be *locative*, geolocated in a predetermined space, or *site-specific*, integrated into a place. The integration offers three dimensions for consideration, quoted from [14] below:

- *Esthetic*: the way in which the experience engages or reconfigures the user's perceptual relationship to the environment.
- *Performative*: the way the experience brings us into an active or possibly interactive relationship with that environment.
- *Social*: the capacity of the experience to take the user beyond individual esthetic or performative responses and to connect her to others toward achieving a variety of individual and collective goals.

For Bekele [15], MR is defined through its capacity to create a real-virtual environment that enhances our perception of both environments. Users, virtual components, and the real world may all interact in this environment, establishing

a user-reality-virtuality interaction and relationship space. Here, MR is discerned from AR through the equality of the real and virtual, where both environments benefit from each other's elements.

So far, we've discussed Mixed Reality as it pertains to the conceptual frameworks that seek to define it. There is no single definition of MR, and understanding is generally based on context. For the purpose of this paper, we offer the following working definition: *MR is a real environment augmented with virtual objects (in effect, AR) distinguished by an increased emphasis on user and virtual object interaction with space.*

3 Related Work

In the following section, we explore existing applications and performances of XR technologies as well as traditional DMIs. We offer a thematic approach, exploring early XR experiments, XR systems in live performance, education, multi-modal musical interfaces, and collaborative music-making.

3.1 Early Experiments

We return to Lanier's work on the *The Sound of One Hand* [1]. Each performance was entirely improvisational, creating a unique experience for both the performer and audience. Lanier described the instruments as "somewhat autonomous", noting that they "occasionally fight back"[4]. The most ergonomically complex of these instruments was the *Cybersax*, which allowed the musician to play a melody over a large range, while at the same time controlling the overall mix of the music, as well as parameters, timbre, volume, and placement of tone. Lanier notes that this purpose of this was to allow the performer to play music in an intensely gestural style.

In 2005, Mäki-Patola *et al.* introduce and analyze four gesture-controlled musical instruments: the *Virtual Xylophone*, *Gestural FM Synthesizer*, *Virtual Membrane*, and *Virtual Air Guitar* [16]. The *Gestural FM Synethesizer* was created as an evolution of the theremin, where pitch and amplitude may be adjusted by moving the right hand up and down and opening and closing the fingers. In addition, the instrument provides a visual representation of a musical scale in the form of a vertical piano. Pitch is indicated on the piano by a thin line extruded from the musician's hand. The authors found that the inclusion of visual feedback made it simpler for users to locate certain notes, compared to the original Theremin. The aim of the *Virtual Air Guitar* was to develop an actually playable version of the experience, where the distance between the user's hands is used to determine pitch, and plucking is accomplished by moving the right hand in a strumming motion. Slides are created by gliding the left hand along the imagined guitar neck, while vibrato is created by shaking the left hand. Of all the four instruments, the Virtual Air Guitar was the most popular, which the authors ascribe to the "cool" nature of the instrument, as well as its intentional design that allowed for even inexperienced performers to translate their movement into decent sound.

3.2 Live Performance

Inspired by Lanier, *Connexion*[5] is an audiovisual music performance for VR that shares the scenography of *The Sound of One Hand*, in which the performer's POV

[4] `http://www.jaronlanier.com/vr.html`
[5] `https://www.youtube.com/watch?v=1jyxLFTQKqk`

is projected on a screen behind. In *Connextion*, the performer manipulates a 3D object to control the granular synthesis[6] and spatial positioning of sounds within an auditorium, where the hand movement pans the sound, and hand distance modifies the LFOs and length of the sound grains. Connexion was well received, however during the post-concert discussions, there was a shared interest in observing the performance from within the VR, alongside the artist [17].

Imagined Odyssey is the name of a experimental dance performance by Case Western Reserve's Department of Dance. The performance utilized AR technology, equipping each of the 80 audience members with a Microsoft HoloLens. The experience included a holographic tornado and mysterious trees that grew from the floor (see Figure 3). Comet-like trails of fiery particles were created by dancers' movements [18].

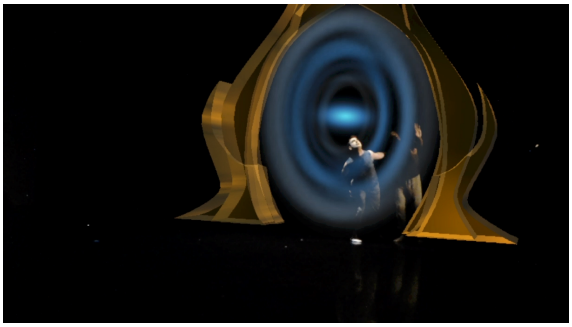


Figure 3: Imagined Odyssey augmented dance performance, video available at: `https://www.youtube.com/watch?v=arqO9vgSOOO`

Imagined Odyssey, although using technology capable of MR expression, only demonstrates the potential for AR. This is namely due to the fact that audience members weren't able to interact with the environment, reducing the use of the HoloLens to a see-through monitor-based system. However, another performance titled *ALIVEmusic* (Augmented Live music performance using Immersive Visualisation and Emotion), made use of both MR technology and audience interaction [19]. In this performance, audio-reactive computer-generated visuals were viewed through either a Microsoft HoloLens or mobile device. A laptop instrument was used to trigger sample sounds which influenced the visuals through various musical properties. Concurrently, a camera-based emotion sensor tracked participants' facial expressions and predicted their expressed emotion which wasmapped to emoticons on a screen. Unlike *Imagined Odyssey*, participants were free to move around the performance space, though similarly, there was no interaction with the virtual objects. A key aspect of this study was to investigate if participants felt more connected to musical expression following the augmented performance. The

[6]Granular Synthesis is a method by which sounds are broken into tiny grains which are then redistributed and reorganised to form other sounds

study found that the HoloLens' narrow field of view, the laptop instrument's lack of expressiveness, and the discomfort of handling a mobile device were the largest barriers to immersion.

Audience participation through Augmentation was explored in *con i piedi per terra*, the first participative musical performance based on the Augmented Stage concept [20]. Audience members were given the opportunity to contribute to the sound of a performance by manipulating AR virtual objects through the touch-screen of their mobile device. The audio parameters were mapped to features of the objects, such as the position of an object to the distortion of a bass synthesizer.

Figure 4: *con i piedi per terra* stage [20]

This concept was also explored in *Virtual_Real*, a multimodal hybrid reality performance characterized by three qualities: the co-existence on the stage of a human element and a machine element, influence sound and music through the manipulation of the graphic environment, and the active role the audience has during the performance [21]. In this setup, the performer and audience members were equipped with reflective marker combined with an IR tracking system for optical motion capture. A screen was used to display stereoscopic projections to the audience through shutter glasses. The markers allowed the artist to control audio effects by touching and morphing unnatural shapes.

Hamilton examines examine how conventional bowed string instruments' performance gestures and mechanics may be translated into a non-physical application through VR. Using *Coretet*, musicians have access to a networked performance environment that supports and presents a traditional four-person string quartet [22].

3.3 Musical Pedagogy

An interesting aspect of study is the use of XR musical systems for learning. Johnson and Tzanetakis present *VRMin*, a mobile XR application that augments a physical theremin with an immersive virtual environment for real time computer assisted tutoring. The setup included a Moog Therimini, as well as use of the Google Daydream platform and a Google Pixel XL [23].

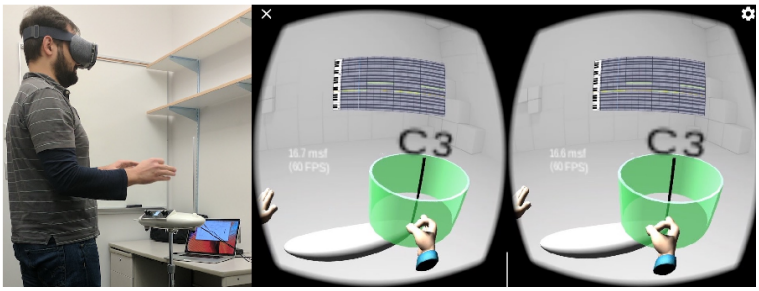


Figure 5: VRMin [23]

This effort is succeeded by the *MR:emin*, which utilizes a Samsung Odyssey HMD, with 6 DoF inside-out tracking, on the Windows Mixed Reality (WMR) platform. The system was used to evaluate traditional music learning with two virtual learning environments, an immersive one and a non-immersive one. A key finding from this study concluded that the majority of participants (n = 30) preferred an immersive learning environment. However, the lack of spatial awareness between the real world and the virtual world presented as a challenge. Though participants hands were visible, this is speculated as the result participants' bodies not being represented in the virtual world [24].

There is also large body of research on XR-based musical pedagogy systems that aim to teach the Piano. One such example is HoloKeys [25], which runs on a head-mounted display (HMD) worn by the user while seated in front of a physical piano. Note objects are produced in the distance and then begin to move towards the specific keys. The note should be sounded as soon as the virtual object hits the actual key. In [26] users at the piano may interact with interactive lessons, view virtual hand demonstrations, see and hear sample improvisations, and perform their own solos and accompaniment with the help of AR-projected virtual musicians. The tool seeks to be both interesting and practical in teaching fundamental musical principles.

3.4 Multi-modal Interaction

We have already examined musical XR systems that use gestural styles, such as the *Cybersax* and *Virtual Air Guitar*. Next we look at systems that utilize other

input methods. *EyeHarp* is a gaze-controlled musical instrument that uses Dwell time as the primary method of input (not an XR system) [27]. Here, a selection occurs when an item is focused on for an extended period of time that exceeds a predefined time-threshold. Alternatives include blink-based selection, where to choose an item, the user stares at it and blinks, and gaze gestures where by producing a series of "eye strokes" in the correct order, the user may execute a command. The pEYE menu, the Step Sequencer, and the Arpeggiator make up the three main levels of the EyeHarp interface. Both of the latter are used to create a rhythmic and harmonic foundation for the composition, while the pEYE menu is used to play melodies and change chords in real time (see Fig 6).

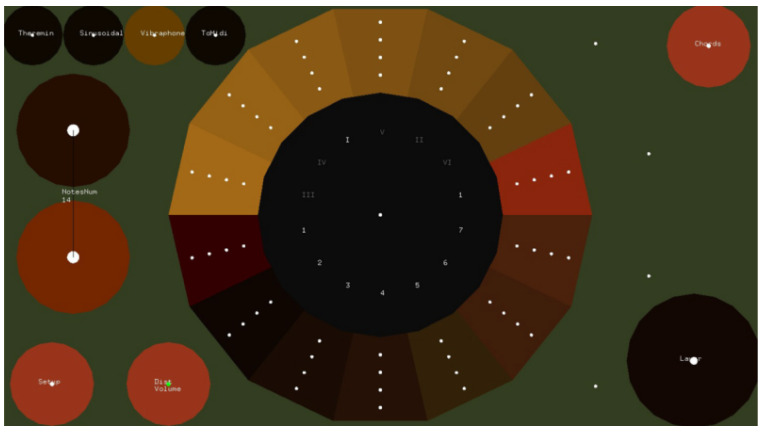


Figure 6: EyeHarp [27]

Davanzo and Avanzini examine a selection of hands-free accessible DMIs, one of which is the EyeHarp. They also reference a project titled *Eye Play The Piano*, in which eye-tracking through an HMD allows for gaze pointing to determine the selection of both piano notes and chord sets, and blinking is mapped to trigger note on and off. Visual feedback is provided through the HMD, through the signal of key selection [28].

3.5 Collaborative Music-making

Next, we look at two collaborative systems. *Alive* is an instrument that allows several users to collaborate on the development of auditory and visual behaviors of virtual agents while engaged in a virtual environment via spatialized audio and stereoscopic projection [29]. Its websocket-based interface operates in any modern web browser, where users are able to author, invoke, remix, and copy fragments throughout the duration of a performance. To facilitate collaboration, each performer sees and modifies the same live document of code.

Figure 7: Alive: AlloSphere [29]

Shoggoth is a network music program for real-time distributed group performance. The system allows users to reshape polymorphic terrains to create generative music in collaboration [30]. The terrain is initially displayed as a flat grid, which can be manipulated into various shapes using generative processes, such as the Diamond Square Algorithm and Cellular Automata. Each terrain produces a synth that is defined by its own shape, that is a wave terrain oscillator that reads the 2D height map. The terrain affects the triggering of synth instances, modulates synth parameters, and determines the synthesizers' tone. User-defined sequences are also mapped on the terrain. Players are represented by tetrahedrons, with networked position and rotation, allowing performers to locate one another. While not an XR system, Shoggoth is one of the primary sources of inspiration for our novel MRMI discussed further in Section 5.

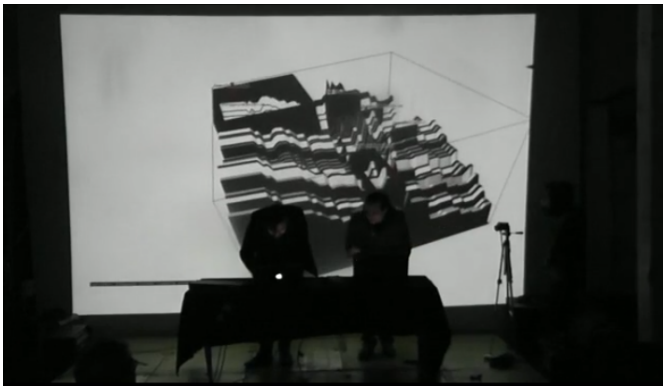


Figure 8: A live performance with Shoggoth, video available at: `https://vimeo.com/94046155`

4 Analysis and Design Framework

This section details the first of our two-part contribution. We review existing research, discuss the affordances of MR, describe our framework with application to select case studies, and assess it through expert interview.

4.1 Existing Frameworks

To reach our own definition of an MRMI, we offer a brief history of virtual instrumentation. A Virtual Musical Instrument (VMI) is a musical instrument with a virtual control surface that is influenced by the physical world in some way [31]. With the emergence of accessible Virtual Reality devices, a subclass of VMIs appeared, known as VRMIs. These include a computer-generated visual component mediated by an HMD or other types of immersive visualization interfaces [32].

There exists a limited selection of frameworks for evaluating VRMIs. In 2016, Sefarin *et al.* outlined a set of nine principles for designing VRMIs [32].

1. Design for Feedback and Mapping
2. Reduce latency
3. Prevent cybersickness
4. Make use of Existing Skills
5. Consider both Natural and "Magical" Interaction
6. Consider Display Ergonomics
7. Create a Sense of Presence
8. Represent the Player's Body
9. Make the Experience Social

Based on these principles, a three-layered evaluation framework was proposed. The first layer deals with interaction modalities, such as input and output, as well as perceptual integration and mapping dependent on users' sensorimotor and cognitive capacities. The second layer is a VR-specific layer that caters to cybersickness, virtual body ownership and representation, and presence. Finally, the third layer tries to assess the objectives, methods, and experiences of users.

Another direction for evaluation pertains to scenography, the study of a performance's visual, experiential, and spatial composition. In [33] Berthaut *et al.* redefine scenographic considerations as they pertain to performance setups that include VR systems. They refer to these systems as Immersive Virtual Musical Instruments (IVMIs), which rely on the depiction of sound processes and parameters as 3D objects in a Virtual Environment (VE). They offer six dimensions for the scenography of IVMIs, quoted from [33] below.

1. *Musician Immersion*: how well the musician(s) can perceive the VE and therefore the instrument
2. *Audience Visibility*: how well the musician(s) can perceive the audience
3. *Audience Immersion*: how well the audience perceives the VE and therefore the instrument
4. *Musician Visibility*: how well the audience can perceive the musician(s)
5. *Gestures Continuity*: how the musical gestures performed by musicians in the physical space are connected to the graphic feedback of the instrument's metaphor, as perceived by the audience
6. *From Virtual to Physical*: how the virtual and physical spaces are merged

4.2 From VR to MR

To further differentiate our class of proposed MRMIs from VRMIs, we emphasize MR as an experience where the user is *not* fully immersed in a virtual environment. As such, many of the themes from literature on VRMIs and IVMIs require revision. We explain three such themes as follows.

4.2.1 Cybersickness

The term cybersickness was proposed by McCauley and Sharkey in 1992, describing the interim side-effects caused by virtual reality immersion [34]. This is sometimes referred to as "simulator sickness", a term that was initially coined to describe the effects induced by simulators but has since been adapted to non-simulator virtual experiences. One study found that the total severity of cybersickness was approximately three times greater than that of simulator sickness [35]. There is currently no agreement on which terminology should be used in respect to modern VR technology [3]; in general, this paper will use the term cybersickness going forward.

Cybersickness symptoms include disorientation, headaches, sweating, eye strain, and nausea. The most commonly accepted explanation for this phenomenon is that cybersickness occurs as a result of conflicting information from the visual and vestibular senses; this is termed Sensory Conflict Theory. Display latency, flicker, calibration, and ergonomics are all thought to influence cybersickness [36]. Susceptibility to cybersickness may also be affected by individual differences such as gaming experience and sex [3][36][37].

As such, VRMI developers are encouraged to minimize accelerations and decelerations, should the user need to move virtually while being physically stationary [32]. Furthermore, as AR and MR systems present content in a more realistic and embodied context, such conflicting factors may be considered negligible. In [38], it is noted that the inclusion of real-world visual references maintains the observers' regular stability conditions, hence significantly lowering sickness effects.

In [39], the use of the simulation sickness questionnaire (SSQ), a standard in research [40], found that there is almost no simulator sickness when using the Microsoft HoloLens. Though, the study did not report any differences based on sex. Therefore, while we do not dismiss the potential for cybersickness in MR, we argue that it is less cause for concern and not a notable factor NIME practitioners should consider when developing XR instrumentation.

4.2.2 The Player's Body

In VR systems, individuals are unable to see their own body portrayed in the virtual environment unless the real body is monitored and mapped to a virtual representation [32]. This concept, known as virtual body ownership (VBO), is a key area of study for both VR researchers and psychologists alike, as it contributes to the wider field of body ownership illusion. In recent years, VR has been used to explore virtual body ownership and agency, under the term virtual embodiment [41][42]. Where agency refers to the notion that a person recognizes themselves as the cause of the actions and movements of that body.

However, virtual body ownership can have transient effects on user attitudes and behavior in the context of musical performativity due to differences between the real and virtual body [42]. This phenomenon is often absent from MR experiences, as the user's body is typically visible, and a key component of interaction. While there may be instances in which a virtual body is desired, such as for telepresence purposes [43], representing the player's *own* body should be of less concern.

4.2.3 Presence

Presence, the sense of "being there", is a phenomenon of human experience that occurs in the context of technologically mediated perception. It can be defined as the combination of two orthogonal components: place illusion and plausibility [44]. The former refers to the quality of having a sensation of being in a real place, while the latter refers to the illusion that the scenario being depicted is actually occurring. The Igroup Presence Questionnaire (IPQ) [45] is a tool designed to measure a user's sense of presence in a virtual environment. Research on measuring presence in augmented or mixed reality environments is still exploratory [46], though work has been done to create a standardized measurement, such as the Mixed Reality Experience Questionnaire (MREQ) [46].

In order to create a convincing sense of presence, perceptual consistency is key [47]. In VR, this is often a challenge, as providing a unified sensory model can be difficult, especially when motion is involved. As such, research has shown an inverse relationship between simulation sickness and presence [37], whereby greater presence can draw attention away from sensory conflict, and less sensory conflict can create greater presence. Depth is another factor of presence. In VR,

depth perception was underestimated, while no underestimation was observed in AR [48]. It is simple to create some sensation of depth, while the difficulty remains in creating an accurate sense of depth [47].

Another component of presence is visual fidelity (reproduction fidelity) [5]. For both the physical and virtual realms, consistent visual quality is critical, as measured by resolution, framerate, and latency for example [46]. VR hardware often has the benefit of providing high-fidelity visuals, by utilizing PC processing power through tethered connections, and a wider field-of-view [49]. Thus, it is possible in VR to render all three domains (environment, objects, people) with the same fidelity within one virtual environment [41]. However, applying the same approaches to three-dimensional MR is difficult and computationally expensive. Though research has found that similar results can be achieved in MR systems with low visual fidelity, by decreasing the realism of one or both visual realms, real and virtual, to achieve visual coherence [46].

This notion of presence, known as presence-as-feeling, has significant implications for musical performance. Flow cannot be experienced without a sense of presence, as it requires the musician to be entirely immersed in the created musical reality whereby the musical instrument has disappeared from consciousness [50]. As a result, the musical instrument is unconsciously perceived as an extension of the self culminating in the synthesis of musician and musical instrument.

4.3 Proposed Dimensions

As we have discussed, a technocentric understanding of MR may impede the primacy of user experience, and limit future exploration. Thus, we broadly define an MRMI as *an embodied system for expressive musical performance, characterized by the relationships between the performer, the virtual, and the physical environment.*

Following this definition, our framework is based on three *interconnected* dimensions: embodiment, magicality, and relationships. These dimensions were inspired by existing frameworks for virtual instrumentation and chosen as we feel they are broad enough to encapsulate all elements of MRMI design, but still useful for analysis. We provide guiding questions for each dimension and offer relevant examples based on current technologies. The first dimension is applicable to both VRMIs and MRMIs, while the last two consider affordances specific to MRMIs.

4.3.1 Embodiment

This dimension considers the way in which entities are mapped to sonic parameters, how they are interfaced, and the feedback offered to users.

- How does the representation of objects affect musical expression?
- How diverse is the range of embodied musical output?
- What feedback could create a better understanding of musical expression?

- How could alternative input methods increase expressivity?

Physical instruments often provide multi-modal feedback, with varying configurations of auditory, visual, and haptic. As such, adding tactile feedback to computationally mediated systems can improve the music playing and learning experience significantly [51]. Some options include ultrasound vibrations for mid-air haptic feedback [51] or the use of electrical muscle stimulation for force feedback [52].

Thoughtful design for mapping and feedback is essential for perceptual consistency, where all sensory signals feed a single mental model of the world [47]. In turn, the level of bodily engagement, as measured by the degree to which action and perception coincide, helps determine the quality of the musician's experience [53]. Many VRMIs are single-process instruments, meaning they can only control one synthesis or effect process at a time [54]. The fundamental advantage of graphical musical interfaces, on the other hand, is the opportunity for multi-process control with visual feedback. [54] explores the use of 3D reactive widgets, which allow both manipulation and visualization of a musical process, whereby its graphical parameters are bidirectionally connected to the parameters of the associated musical process. Techniques for manipulation of elements in a virtual environment include spatial transformations (rotation, scaling, translation), structure manipulation, and material manipulation [55].

Though while virtuality allows for additional control dimensions, this does not necessarily mean that an instrument's expressivity is correlated to the number of controls or degrees of freedom (DoF) [56]. In [57], researchers discovered that adding a control dimension to an instrument (1 DoF vs 2 DoF) actually lowered the exploration of hidden affordances, and that participants in the 1 DoF group felt there were more features remaining to investigate than those in the 2 DoF group. From this, they concluded that the development of diverse playing styles is a common feature of highly constrained instruments. Hunt and Kirk [58] examine various strategies for mapping human gestures onto synthesis parameters for live performance. They find that "real-time control can be enhanced by the multiparametric interface" and "mappings that are not one-to-one are more engaging for users".

4.3.2 Magicality

This dimension considers both the "magicality" and "naturality" of interaction.

- How is this interaction made possible by virtuality?
- How is this interaction contributing to the relationship between gesture and result?
- What natural constraints are being observed?
- What metaphor is being used?

With acoustic music, it is physically evident how the sound was produced, with close to a one-to-one relationship between gesture and result [59]. Now with the power of computers as the intermediary between our physical body and the sound production, we may "go so far beyond the usual cause-and-effect relationship between performer and instrument that it seems like magic. Magic is great; too much magic is fatal" [59]. In the context of VRMIs, either an interaction or an instrument can be considered magical if it is not constrained by real-world restrictions such as those imposed by physical laws, human anatomy, or the present state of technical advancement. Conversely, interactions and instruments qualify as natural if they adhere to real-world conditions [32].

We adopt this idea for MRMIs, yet pose that greater naturality is an affordance provided by MR systems, as digital content is presented to the human perceptual system through direct integration into the physical surroundings [47]. This natural baseline should encourage MRMI developers to explore magicality, attempting various combinations of magical and natural interactions. For example, in [24], a physical theremin is augmented with an immersive learning environment, providing real-time visual instruction and feedback for note placement.

We may also consider magicality with respect to *transparency*, which is defined as "the psychophysiological distance, in the minds of the player and the audience, between the input and output of a device mapping" [60]. Fels *et al.* argue that transparency is a predictor of expressivity, and through metaphor, transparency increases. Metaphor in this instance is used to restrict and define the mapping of a new device, transforming it from an opaque mapping to a transparent one. In Sound Sculpting, the metaphor of sculpting clay was applied to change the shape of a virtual object, which in turn affected the parameters of an FM synthesizer. The study found that certain aspects of the mapping were self-explanatory, while others were obscured by the metaphor, emphasizing the importance of selecting a metaphor that is compatible with the input and output interfaces. Here, we can assume that magicality is inversely related to transparency.

However, others propose that a lack of transparency, and therefore, magicality can become an asset rather than a hindrance [61]. This is based on the idea that novel instruments seek to be both a tool to perform music and part of the musical composition itself, whereby each new instrument represents a unique interpretation of the relationship between action and sound, and comprehending this interpretation may be just as artistically fulfilling as listening to the music itself.

4.3.3 Relationships

This dimension considers the network of relationships between all entities, user, physical and virtual. Rather than focusing on technical aspects, this dimension is centered on user experience, including multi-user collaboration.

- What connection do I have to my physical environment?
- What connection do I have to my virtual environment?
- What connection do I have to the other performers (if applicable)?
- What connection do I have to my audience (if applicable)?

In VR applications, the user is transported to some location, immersed in a synthetic reality. Thus, these systems typically do not consider the user's physical location, but rather, require that it is empty, or at least free from obstruction. A recurring theme in existing research references the limitations of VRMIs due to the occlusive properties of HMDs (non-see-through) [62] [32] [33]. In relation to scenography, HMD usage prevents the audience from being seen by the musician, resulting in a total absence of audience visibility. Many VRMIs are also alienating as they inhibit the development of relationships between the performer(s), the audience, and the surrounding space.

Whereas in MR systems, location plays a vital role in the context of the spatial position of both virtual objects and user(s), whether intentional or not. MR enables new forms of storytelling by allowing virtual content to be meaningfully linked to specific locations, whether they be places, people, or objects [63][64]. While not directly related to MRMIs, [65] describes an interesting application of this affordance: an audio-based MR experience that invites visitors to learn about the culturally and personally significant events of a cemetery's departed residents. Though this opens discussion on the potential for alternative types of musical augmentation.

In [66], musical objects were represented by simple 3D shapes, with mappings left as esthetic choices made by the composer. Options for interaction included looking at an object to change its trajectory, crouching or standing up to shift the cutting frequency of a low-pass filter, and traversing in space to activate audio effects. This study offers an exciting taste of what musical interaction in MR could look like, where the relationship between body and space is explored as part of the performance.

Lastly, we consider collaboration in music-making. MR is an ideal host for collaborative interfaces because it addresses two primary concerns in computer-supported cooperative work: seamlessness and enhancing reality [64]. When co-located, users can see each other's facial expressions, gestures, and body language, increasing the communication bandwidth. This is significant since it is often the group atmosphere and the establishment of synergistic interactions between players, rather than the interface itself, that leads to positive communal experiences in music-making. MR systems can offer independence and individuality, where each user controls their own independent perspective, and displayed data can be unique to each user [67].

4.4 Case Studies

This framework is not intended to distinguish a "good" or "bad" MRMI, but rather, provide dimensions for their design, discussion, and analysis. There is ample opportunity for further exploration in this emerging field, especially with the limited selection of current MRMIs, even under the broad definition we provide. While hardware is currently expensive, low-cost alternatives[7][8] are promising options for these early stages. We will now apply this framework to three existing applications we consider to be MRMIs.

Augmented Groove is a musical interface that explores the potential for augmented reality, 3D interfaces, and tactile, tangible interactivity in multimedia musical performance [68]. Users can collaboratively create music by manipulating physical cards on a table, which are mapped to sonic properties, such as timbre, pitch, rhythm, distortion, and reverb. Users wearing (see-through) HMDs can view 3D virtual images attached to the cards, the forms, colors, and dynamics of which correspond to musical elements. **Embodiment** is certainly a critical dimension of this system, as the music takes on the form of a solid, tactile entity that can be handled and seen as part of the physical world. The input of the system is dictated by its physical, tangible interface, where all the user needs to do is pick up and move the cards. This bleeds into the balance between **naturality** and **magicality**, as the interaction with the cards is simple and intuitive, where the relationship between gesture and result is preserved by direct mapping. This contributes to the **magicality**, both in the improvisational nature of the system leading to uncertainty of the resulting sound, and the excitement of a new relationship (between card and music) made possible through technology. The **relationship** dimension is equally well-explored, as users can see the physical world, virtual objects, and each other, interacting and passing around sequences. The importance of collaboration permeates not only through the relationships between the performers, but the relationships between the physical and virtual objects. Connections are formed between performers as they collectively author and improvise music.

A Very Real Looper (AVRL) is an audio-only virtual reality interface inside which a performer controls musical sounds and sequences through gesture and full-body movement (see 9) [62]. The system maps virtual musical sounds onto tangible items in the real world using two VR sensors and the Unity game engine. These sounds may be triggered, repeated, acoustically altered, or relocated in space using two hand-held VR controllers. The result of this is a system that enables expressive and embodied musical interactions. Unlike its original categorization as a non-visual VR interface, we consider this application to be an MRMI,

[7] `https://www.zappar.com/zapbox/`
[8] `https://www.lynx-r.com/`

and one that shares many similarities to Augmented Groove. As such, **embodiment** is once again a key component of the system, as sounds are mapped onto physical objects. When the system detects a collision between the virtual objects and hand-held controllers, a musical sample, a MIDI sequence, or a specific MIDI note or chord is triggered. This audio feedback is aided by haptic (vibrational) feedback from the controller. Thus, these interactions create a system where the performer is physically colliding with music. However, left to be uncovered is any meaning behind the object and sound. Does a sound mapped to a rock bear any rock-like features? The **relationship** dimension is one of particular interest in this work, as the performer is able to develop a relationship with the audience and surrounding space. Furthermore, the performance is site-specific, where the positioning of objects can dictate the direction of the resulting sound, creating unique experiences of both sound and interaction based on placement. For instance, a virtual object placed on a light fixture would require the performer to jump or throw the controller to trigger that sound. The balance between **naturality** and **magicality** may be different for the performer and audience. Assuming the performer is initially aware of the mappings between physical objects and sound, it is left to be discovered by the audience over time.

Figure 9: Technical breakdown of *AVRL*, video available at `https://vimeo.com/288778622`

Touching Light is a framework for the facilitation of Music-Making in Mixed Reality [69]. In this performance, a Microsoft HoloLens was used to augment live music-making through a series of distinct movements. In the first movement, Simplicity, a holographic mixer is used to modify the audio parameters of accom-

paniment tracks. This exploration is more **natural** than **magical**, as it is simply a virtual representation of a physical sound mixer. Besides its inherently holographic nature, which provides scaling, rotation and repositioning, there is nothing that is added or made possible by virtualization. **Embodiment** is interesting, as both the use of physical instruments and the virtual mixer levels guide the musical expression. Here, the virtual environment is interfaced through gesture-based controls, where the performer performs a pinch gesture to select and slide each fader. Though the mapping is fairly simple, with slider values controlling the volume of ten distinct tracks. In the second movement, *Soliloquy*, a virtual carousel of images rotates around the performer, serving as a critical element of the score that is notably not possible in traditional Western notation. The images themselves hold no agency in the music-making process, but rather exist to inspire the improvisation of the performer. In this sense, the **relationships** between the performers and the images impact the resulting musical content, as an indirect mapping. In the third and final movement, *Synecdoche*, three holographic cubes emerge in the surroundings as little music-making satellites in space, unbound by gravity yet present onstage with the artist. These cubes collide with the real environment, ricocheting and rebounding, turning real. This movement offers the most interesting balance between **naturality** and **magicality**, as these floating cubes defy the laws of gravity, yet collide in such a way that is perceptually consistent. These interactions are also unique to MR, and offer a glimpse of what is made possible by virtuality. Lastly, there are several **relationships** in play, such as the connection between the performer, the physical instruments, and virtual entities.

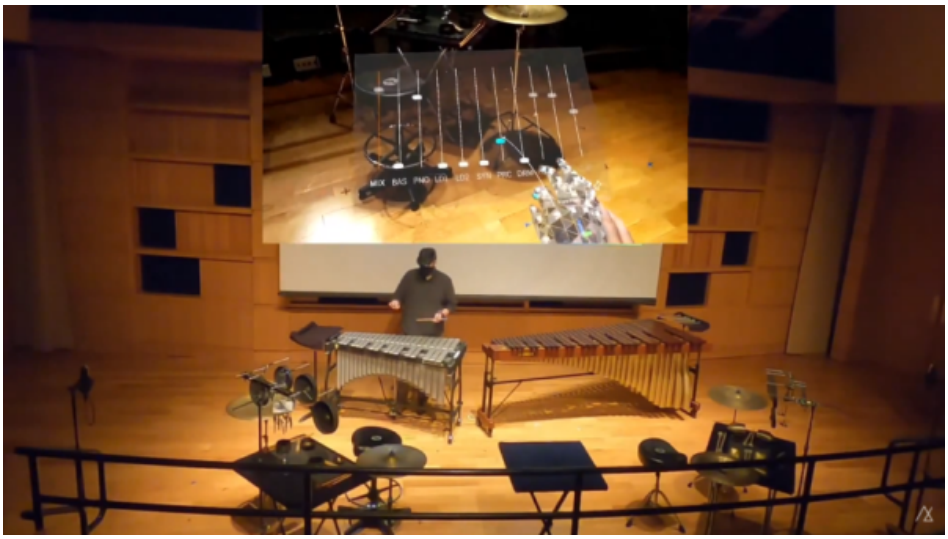


Figure 10: A live performance of Touching Light, video available at: `https://www.youtube.com/watch?v=4UeQApWRWlM`

4.5 Evaluation

This framework was accepted to the NIME 2022 conference after being peer-reviewed by four reviewers and one meta-reviewer. Based on their feedback, the framework was improved and refined to its current state.

4.5.1 Reviewer Feedback

One of the first changes was expanding related work solely when it applied to musical interaction. Specifically, we removed some of the more general discussion on cybersickness, such as existing VR-related mitigation research, and only kept the content we felt was essential to explain why MR might be a better choice for NIME practitioners. Likewise, presence was modified to include only the necessary concepts required to convey how presence is correlated to musical interaction and how MR may achieve this.

The dimensions themselves also underwent significant change. Text was added to the beginning of the framework to explain how and why these specific dimensions were chosen. Then, each dimension was reworked in several ways. First, their description was improved to address how they relate to concepts presented in the background on MR and to aspects of DMI design. Second, additional context was added to explain how they may help in the design and analysis of an MRMI. In embodiment, for example, the relationship between control and musical expressivity was explored. This also included a discussion of its relevance to the concept of transparency and metaphors. For relationships, this was accomplished through discussion of multi-user performance and spatial audio. Magicality was improved to address its implications in term of design choices and the impact it has for performers and audiences.

It is also important to note that there was formerly a 'Modalities' dimension, however it was significantly weaker than the others and certainly the least defined. To address this, we decided to combine Modalities and Embodiment, reducing the framework to three better-defined dimensions. Research on musical gestures, for instance, was then better positioned by the new and improved embodiment dimension.

4.5.2 Expert Interviews

We now further evaluate this framework through the results of four semi-structured expert interviews: two from academia and two from industry. The experts selected ranged from four to ten years of experience working with XR technologies and comprised of:

(E_1) an Assistant Professor whose research surrounds musical performance and immersive media.
(E_2) a PhD Student studying the intersection of music & AI.

(E_3) a Senior Technical Designer working on consumer Mixed Reality experiences.
(E_4) an Audio Designer working on consumer Mixed Reality experiences.

The following questions were used to stimulate initial discussion on the framework.

1. Are there important aspects of MR design that the framework doesn't capture?
2. What other issues do you see with the framework?
3. How could this framework be improved?
4. Do you have any additional comments you'd like to share?

Following completion of the interviews, the recordings were transcribed, and pertinent themes extracted. Points of disagreement, suggestion, and shared opinion were of particular interest and discussed further below.

In general, all of the experts agreed that the paper was valuable, with some emphasizing the importance of developing these kinds of frameworks in the early stages of the field (E_1, E_2). E_3 found that the "application of the framework on case studies led to conclusions that were valid and meaningful", justifying the soundness of the framework itself.

Moving onto the specifics, we look at opinions on the dimensions presented. E_1 found they were "broad enough that they cover quite a bit of ground in regards to musical expression". Concerning the number of dimensions, E_2 argued there was "a case to be made for having as many dimensions as possible for people to choose from", referencing work on the epistemic dimension space for musical devices [70] as an example.

E_1 felt that there could have been more discussion on the overlap between the dimensions. They referred to questions such as, how does relationships interact with embodiment? Similarly, how does embodiment factor into magicality? Is it a negation of embodiment? Does it play off of embodiment? Like the fact that you subvert gravitational forces or rules of physical reality? They also mention that the idea of "spatiality" could have been explored further, especially as it pertains to embodiment and relationships. Here they made the point that for MR experiences we must consider what necessitates this particular platform. Interactivity and spatiality are fundamental elements that justify an instrument in XR, and while the embodiment dimension inherently implies that, there could be more explicit discussion of spatiality as it pertains to MRMIs and musical expression.

Both E_1 and E_4 shared the same criticism of the framework in regards to the application of technocentrism solely when referring to VR. E_1 noted that "the framework should have applied same avoidance of technocentrism to the VR side" and that "justifying MR based on technocentric limitations of VR risks obsolescence". The alternative, pointing out the technical shortcomings of MR, was also considered, such as the limited field-of-view of the HoloLens. On the

same thread, E_4 felt there was too much emphasis on cybersickness, especially when it can be attributed to the limitations of current hardware. In general, the feeling surrounded the necessity for more symmetry in this VR-MR dichotomy. There was further discussion on obsolescence in this space, particularly given the rapid evolution of technology in the XR landscape. E_1 argued that Milgram and Kishino's RV continuum could be regarded as obsolete. In our paper, we describe VR as being synonymous with a VE, situated on the end of the continuum and thereby not within the spectrum of MR. However, they argue that VR is not a virtual environment, as that involves a scenario where all senses are convinced. Furthermore, the definition implies ocularcentrism, which we should particularly attempt to avoid in these musical applications.

Another recurring theme was the need for additional discussion of musical practices in MR (E_2, E_3, E_4). E_2 noted that the framework was presented as UX-heavy, dichotomous to it's sole technocentic counterpart, yet stressed that there are other factors to consider. They noted that in existing NIME research, the music side is too often overlooked, citing motivic analysis from [71] as an example of investigation that brings music back to the forefront of attention. While not an issue with the framework itself per say, they encouraged that music be taken into consideration, even just as a reminder for other people.

E_4 suggested that the framework include aspects of community as well as the nature of pieces to be performed. Specifically, the relationships section, which already mentions the audience-performer relationship, could be expanded to incorporate socio-cultural and performance context. This would be considering why people are gathered, whether for religious reasons, joy, or artistic emotion. Asking questions such as: is there a culture around creating the music? For the instrument itself, is there composed material, is it improvisational? In regards to the case studies offered, they noted there was some ambiguity between an immersive audio-visual installation versus an MRMI, but then questioned the need for an explicit boundary. Towards an audience-performer relationship, E_1 stressed the modern technical limitations that prevent us from bringing a traditional concert audience into XR. While possible, it is not technically feasible to provide headsets to hundreds of people. So until XR becomes very commonplace, it is an inherent question to all these experiences: where is the audience situated?

E_3 wished the embodiment dimension was more musical. As a space where they lack expertise, they worried it may be inaccessible to a non-musical community. To this end, there were several interesting discussions on MR-specific music (E_2). As mentioned earlier, spatiality is certainly one aspect. Others that appeared were "virtual materiality", macrodiversity, and microdiversity [72]. E_4 enjoyed the idea of metaphor, however felt that it was skewed to the visual side, suggesting further expansion sonically. This also led into conversation about the nature of MRMIs, whether they are a representation of a real instrument, a recreation, or an augmented instrument. More questions arose, including what constitutes a novel interface. Specifically, what is borrowed, and how might that affect

the novelty of the instrument?

There was much discourse around the idea of magicality versus naturality (E_2, E_3). E_3 noted that "putting them as mutually exclusive points on a spectrum seemed cumbersome". They were concerned that MR was being labeled as a magic box. They mentioned at one point in time, the electric guitar was magical, the radio, the light switch. Now nothing is magical, just a fixture of the world we live in, an application of physics. The novelty made it magical. They argued that this definition "strikes to the heart of the matter, when MR is as part nature as much as electricity" and suggested verification of the framing of the term. This prompted further conversation, especially pertaining to clarification of the usage of "magical" as something that was intended to describe the interactions that may be had in an MR environment, rather than MR itself. If we go back to the definition: something can be considered magical if it is not constrained by real-world restrictions such as those imposed by physical laws, human anatomy, or the present state of technical advancement, we note that the sheer existence of MR as a medium already asserts itself as a natural system.

E_2 disagreed with the idea presented that acoustic instruments are a one-to-one mapping between gesture and result, referring to the potentially complex inner-workings that may not be immediately apparent to all musicians. Though they agreed with the sentiment, that there is certainly a degree of obscurity that is an order of magnitude more for virtual instruments, they argued that interaction in an MRMI doesn't necessarily need to be understandable, as that limits the way we may think about interface design. They stressed the danger of this way of thought, especially as it risks preventing exploration of more "magical" interaction, and constraining the space in general. Justification was made with the claim that in a piano concert for example, many of the audience cannot see the performers fingers but can still appreciate the music produced. They emphasized that this was one of their bigger concerns in regards to the framework.

There was further discussion on the types of interaction in MRMIs. E_2 described an axis with two sides: systems with discrete objects that can trigger various sounds, and systems that continuously track movement to produce a stream of noise. The former is too predictable, while the latter is too arbitrary. They claim that the "golden middle of those two would be where interesting interfaces pop up".

Overall, the interviews provided valuable insight on the strengths and weaknesses of the framework. To conclude this section, we can derive a preliminary list of action items to be addressed in future work.

- Further discuss the overlap between the dimensions
- Provide more balance by addressing current hardware-specific limitations of MR, and non-hardware-related shortcomings of VR
- Revisit the idea of "Magicality" in MRMI design, verifying its usage, and providing additional examples

- Weave elements of musical practice throughout the framework

5 Wavelength

Now equipped with our MRMI framework, we explore firsthand the potential for expressive performance in MR with the Microsoft HoloLens. This proposed system, *Wavelength*, consists of malleable holographic objects, as polymorphic terrains, which performers reshape and orchestrate to create generative music.

In this section we examine the prominent tools and technologies involved, outline the main features and their implementation, and evaluate the system from a variety of perspectives.

5.1 Microsoft HoloLens

The Microsoft HoloLens is a mixed reality headset developed and manufactured by Microsoft[9]. The HoloLens was the first head-mounted display running the Windows Mixed Reality platform under the Windows 10 Operating System. The pre-production version shipped on March 30, 2016. It's successor, the HoloLens 2, was introduced on November 7, 2019 for a price of $3500.

Figure 11: Microsoft HoloLens 2, accessed from `https://www.microsoft.com/en-us/hololens`

5.1.1 Sensors

The HoloLens works by combining several sensors. The HoloLens 2 has eight cameras on the headset, five of which are used to track its environment (four

[9] `https://www.microsoft.com/en-us/hololens`

environment tracking cameras, one depth camera), two eye tracking cameras, and one regular camera for recording video or taking pictures. The HoloLens 2 provides *inside-out tracking*, which is the ability for the headset to track its environment without the need for external sensors [73]. Mixed reality capture (MRC), an in-built capability of the HoloLens, allows for the first-person view of the merged real and digital worlds to be captured as a photo or video and shared in real-time [74]. Our system depends on all of these sensors, as do the majority of HoloLens applications. In order to record the performance, we specifically rely on MRC.

Table 1: HoloLens 2 Sensors [75]

Head tracking	4 visible light cameras
Eye tracking	2 Infrared (IR) cameras
Depth	1-MP Time-of-Flight depth sensor
Inertial measurement unit (IMU)	Accelerometer, gyroscope, magnetometer
Camera	8-MP stills, 1080p30 video

5.1.2 Spatial Awareness

The HoloLens is constantly tracking its environment and building a 3D model of the surrounding area, a practice known as spatial mapping (see Figure 12). This is an essential feature of the HoloLens, as it enables developers to create compelling MR experiences. With spatial mapping, users can be presented with recognizable real-world behaviors and interactions, for instance, pinning items to the walls, or hiding holograms when the user walks into another room [73]. It also allows for hologram persistence, which refers to the capacity of holograms to remain in their original location following a device reset [73]. In our system, spatial mapping is used to maintain the positions of terrains in the real world.

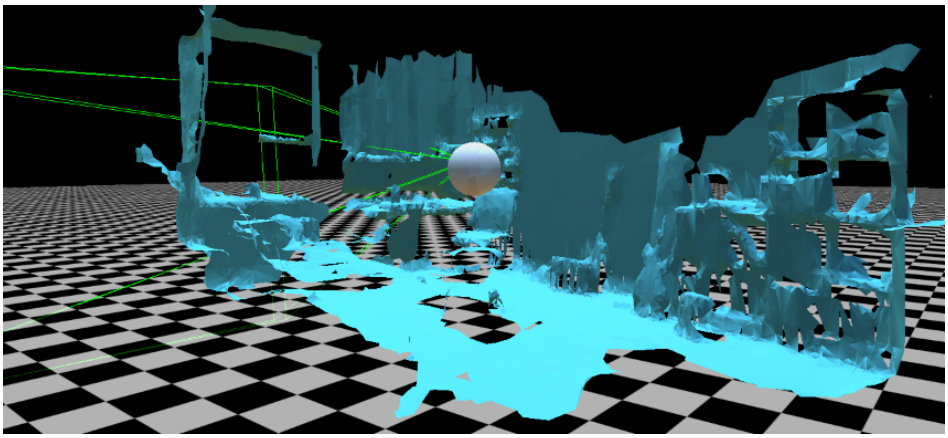


Figure 12: A spatial mapping mesh generated by the HoloLens 2, accessible through the Windows Device Portal

5.1.3 Interaction Models

Unlike most modern VR headsets, the HoloLens doesn't come with hand-held controllers. Instead, a hand-tracking and gesture system is used to define interactions with the virtual environment. Common gestures include Air Tap, Pinch, Drag, and Hand Ray (see Figure 13). By allowing developers to utilize information about what the user is looking at, the HoloLens 2 enables for a new degree of context and human understanding within the holographic experience. The role of gaze in the HoloLens is to that of a mouse pointer on a PC [73]. Coupled with voice input, the HoloLens can offer an entirely hands-free experience. The *"see it, say it"* model allows users to simply read the name of a button to activate it. Our system utilizes all of these interaction models, with heavy reliance on the common gestures, described further in the next section.

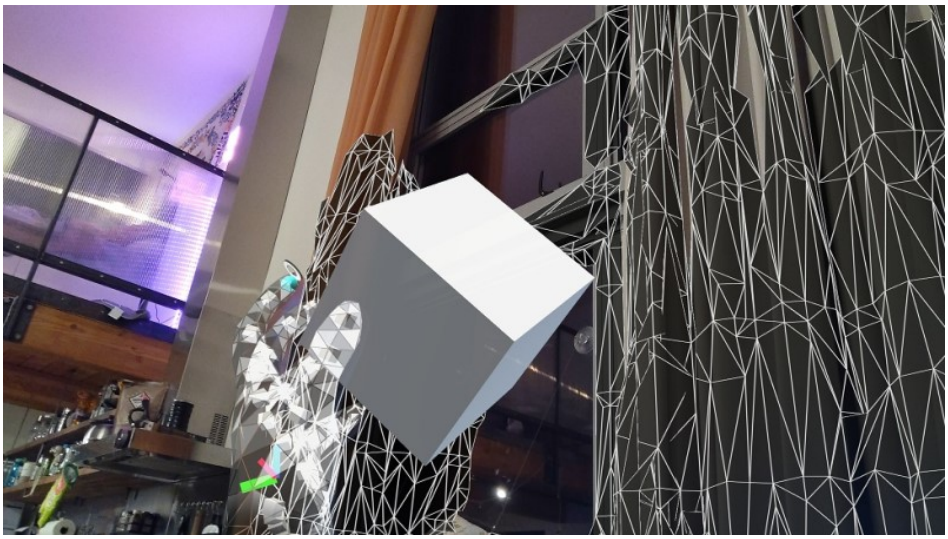


Figure 13: A "near-interaction grabbable" object, selected by the Pinch gesture

5.1.4 Holographic Remoting

Another essential feature is the ability for Holographic Remoting, which allows holographic content to be streamed to a HoloLens in real time without the need to build or deploy a full project [76]. This enables easy debugging during the development process in Unity by allowing developers to play a scene directly from the editor. As such, computation is done on the PC instead of the HoloLens, allowing developers to take advantage of the PC's more powerful resources. Holographic Remoting is especially useful if an app contains high-resolution assets or models that would cause the frame rate to suffer if run directly on the HoloLens. These are known as Holographic Remoting Remote apps. The quality and performance of remoting will differ depending on three factors: holographic experience, PC hardware, and Wi-Fi connection.

5.1.5 Mixed Realty Toolkit (MRTK)

MRTK is a Microsoft-led project that offers a collection of components and services for accelerating the creation of cross-platform MR applications in Unity[10]. Providing a set of reusable components, APIs, and tools, it is an essential part of Mixed Reality development [73]. MRTK comes with several samples, an example of which can be seen in Figure 14. Our system is built on the MRTK framework.

[10] `https://github.com/microsoft/MixedRealityToolkit-Unity`

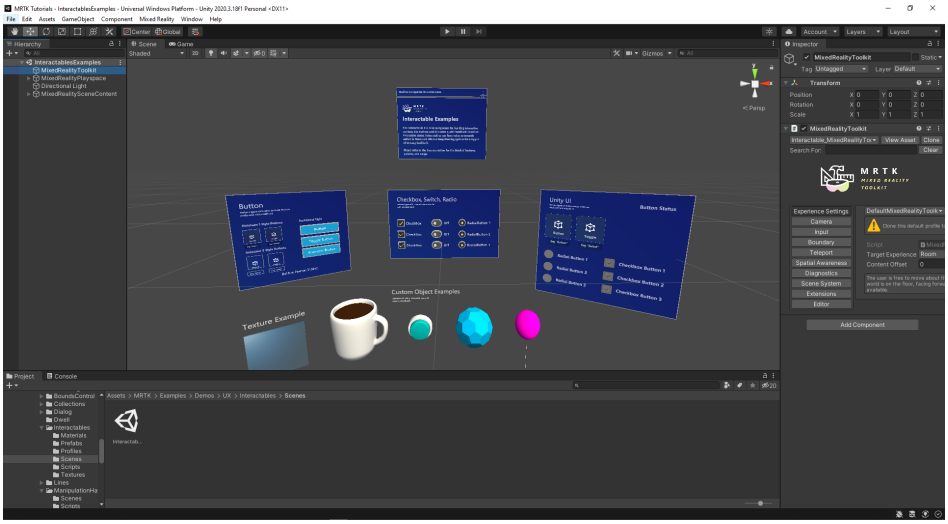


Figure 14: An example scene from the Mixed Reality Toolkit 2.8 in Unity

5.2 Sound Synthesis

Our instrument uses the metaphor of wave terrain synthesis with gestural control of synthesis parameters. We were inspired by existing work, such as earlier Shoggoth performances (section 3.5), and the efforts by Stuart James [77] and Sohejl Zabetian [78] on Wave Terrain instrumentation. We are excited about the potential for dynamic 3D structures, as terrain, to generate both visuals and audio in a unified interaction framework.

We now offer a brief overview of sound synthesis. There are four basic waveforms used in electronic music composition, which are the sine, square, sawtooth, and triangle waves (see Figure 15). Each waveform has a unique sound, and can be used to create different types of music. A sine wave is the simplest waveform, characterized by a smooth, undulating shape. It contains only a single fundamental frequency and no harmonics or overtones, often described as sounding pure or simple. A square wave has sharp, angular edges with a flat top and bottom. It sounds richer and buzzier, due to the presence of harmonics. A triangle wave, as the name suggests, is represented by a triangular shape and contains only odd harmonics. It sounds somewhere in between a square wave and a sine wave. A sawtooth wave has a serrated, saw-like shape, and sounds harsh and clear. It has the richest harmonic content of the four waveforms [79].

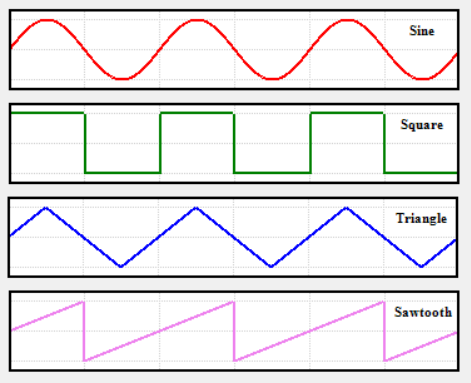


Figure 15: Sine, Square, Triangle, and Sawtooth Waveforms

Additive Synthesis is a method of sound synthesis where the waveform of a sound is created by adding together multiple sine waves. Theoretically, any complex waveform can be closely approximated by summing elementary waveforms—which is the basis for additive synthesis [80].

5.2.1 Signal Modulation

Modulation synthesis is a type of sound synthesis where the timbre of a sound is changed by modulating one or more of its properties. The most common types of modulation synthesis are amplitude modulation (AM) and frequency modulation (FM) [79]. Modulation synthesis can create a wide variety of sounds, from simple vibrato and tremolo effects to complex, evolving sounds.

Amplitude Modulation involves varying the amplitude, or volume, of a sound wave. The wave that is being modulated is referred to as the carrier signal. When a sub-audio signal is used, such as a low-frequency oscillator, the resulting sound is a gradual, undulating effect known as tremolo, where the volume of the sound becomes alternately louder and quieter. Any type of waveform can be used as the control signal, and will have a different effect on the sound. For instance, a sine wave will cause the volume to rise and fall very smoothly, while a triangle wave cause a gradual increase that sharply turns down and decreases. Figure 16 demonstrates analog AM synthesis.

Frequency Modulation involves varying the frequency of a waveform over time. When a sub-audio signal is used (less than 20Hz), it results in a vibrato effect. When the signal used is in the audible frequency range, the resultant signal comprises sidebands of the carrier wave and the rapid undulation of pitch is heard as a change in timbre [81].

Figure 16: AM Synthesis in VCV Rack, `https://vcvrack.com/`

5.2.2 Wavetable Synthesis

Due to the repetitive nature of musical sound waves, an efficient technique for digital representation involves calculating the values for a single cycle of the waveform, and storing these in an array. This array is also known as a wavetable. The process of repeatedly scanning a wavetable in memory is called **table-lookup synthesis**, and is the most fundamental operation of a digital oscillator [80].

For example, assuming the table contains 100 16-bit entries, the oscillator will start at the first entry (phase_index=0), move by an increment to the end of the table (phase_index=99), and then wrap around to the beginning again. The frequency of the sound produced depends on the length of the wavetable and the sampling frequency. If the sampling frequency is 1000 samples per second, and there are 100 numbers in the table, the output frequency is 10 Hz, as $1000/100 = 10$.

$$frequency = \frac{increment \times samplingFrequency}{L}$$

In order to generate different frequencies, the oscillator must resample the wavetable. This is done by skipping values by an increment added to the current phase_index, represented below.

```
var phase_index = (previous_phase + increment) % samples.Length;
var result = amplitude * wavetable[phase_index];
```

However, this logic assumes that the increment is an integer. When in reality, values are often real decimal numbers and therefore incompatible with our integer-based lookup table. An **interpolating oscillator** uses interpolation to determine the value of a wavetable at a specific phase index increment. By interpolating between the entries in the wavetable, the oscillator is able to find the value that exactly corresponds to the specified phase index increment. This process, also known as upsampling, allows for more accurate sound reproduction than non-interpolating oscillators. This also means that smaller wavetables can yield the same audio quality as a larger noninterpolating oscillator.

There are many types of interpolation that can be used, but linear interpolation is usually sufficient [82]. Let vi and vj be values in the wavetable, and we need a sample between them, represented by the real-valued t in the closed unit interval [0, 1]. The linearly interpolated value is given by function:

```
float Lerp(float vi, float vj, float t) {
  return (1 - t) * vi + t * vj;
}
```

Wavetable synthesis is a popular and powerful method of electronic sound production, used commonly in digital synthesizers and DAWs (see Figure 17).

Figure 17: Vital, an application for Wavetable Synthesis from `https://vital.audio/`

5.2.3 Wave Terrain Synthesis

Rich Gold was the first to contemplate utilizing a virtual multidimensional surface to generate audio waves in 1978; he dubbed the surface a Wave Terrain [77]. Wave Terrain Synthesis generates sound using two separate structures: a terrain function and a trajectory signal. The trajectory specifies a set of coordinates that are used to read data from an n-variable terrain function. The most popular Wave Terrain Synthesis techniques employ surfaces represented by functions of two variables, $f(x, y)$, such as in Figure 18.

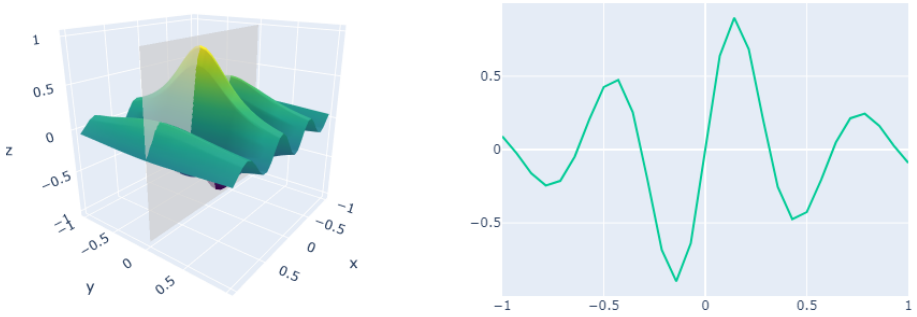


Figure 18: A terrain function $f(x, y) = sin(10x)/(1 + 5(x^2 + y^2))$ with the trajectory signal $y = 0$

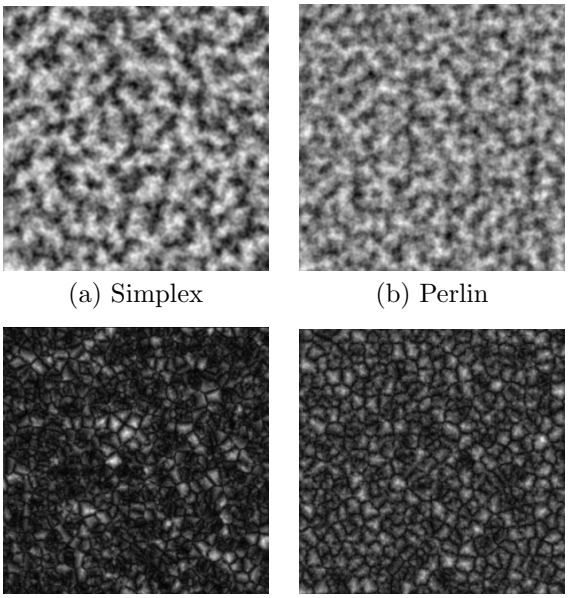
5.3 Features

The system was created specifically for the Microsoft HoloLens 2 using the MRTK 2.8 framework in Unity. We now offer an explanation of some of the features provided by the system.

5.3.1 Procedural Terrain Generation

A crucial aspect of the system is terrain generation. Through a floating panel attached to the terrain, a user is able to select from a wide range of terrain generation algorithms, the majority of which are noise functions (see Figure 19)[11]. The user is also able to select from a sine, triangle, square, and saw wave.

[11] `https://github.com/Scrawk/Procedural-Noise`



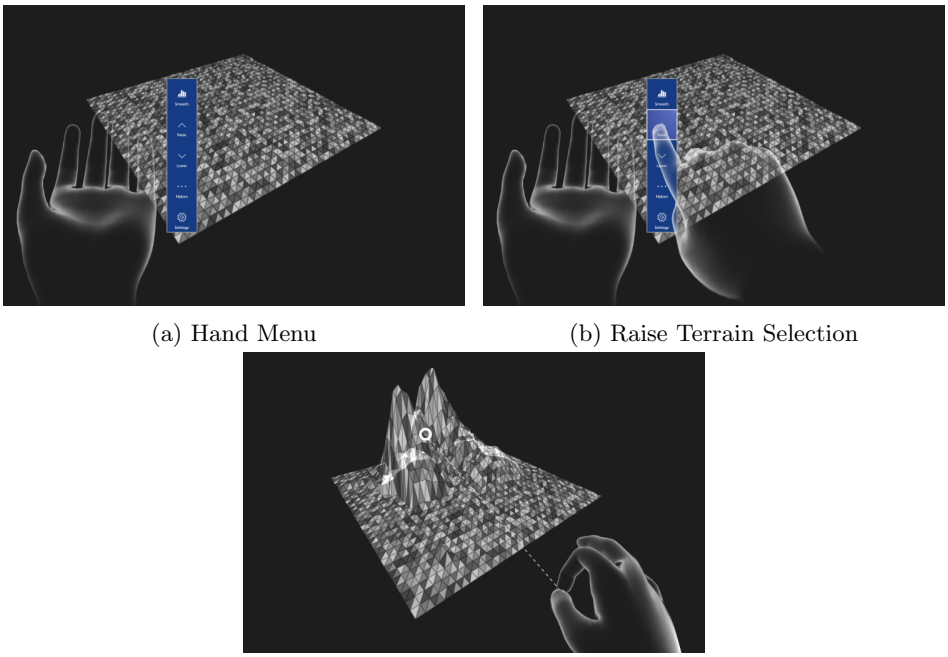
(a) Simplex (b) Perlin

(c) Voronoi (d) Worley

Figure 19: Lattice and point based procedural coherent noise functions, as rendered in Unity by the author

5.3.2 Run-time Mesh (Terrain) Manipulation

Two major affordances provided by the HoloLens are the ability for both hand-tracking and gesture recognition. These enable run-time mesh manipulation, allowing the user to raise, lower, and flatten terrains through gestural input (see Figure 20).



(a) Hand Menu

(b) Raise Terrain Selection

(c) Terrain Manipulation

Figure 20: Far Distance Terrain Manipulation, Mock-up created in Figma by the author

5.3.3 Trajectory Signal

Along with the terrain shape, the trajectory is a critical factor in determining the timbre of the resulting waveform. The system allows for a $z = i$ trajectory function, where i is an integer between 0 and the length of the terrain.

5.3.4 Modulation

Sound modulation is supported through the connection of various parameter ports, such as waveform, amplitude, pitch, frequency cutoff, and traversal. For instance, the waveform output of a trajectory can be linked to the amplitude input of a wave terrain. A number of combinations can be created, with any number of trajectories and terrains.

5.3.5 Audio Effects

The system allows for audio effects and signal processing techniques to be applied to the resulting sound, such as reverb, distortion, echo, and filters (e.g. high-

pass, low-pass) through options in the Hand Menu. The system also provides recognition of a "grab" gesture, which is mapped to the global volume. Basic spatialization exists.

5.3.6 Spectrum Visualizer

The system provides a spectrum visualizer. There are two poles that can be moved along the X-axis, the positions of which correspond to the cutoff frequency of a both Low and High-Pass filter.

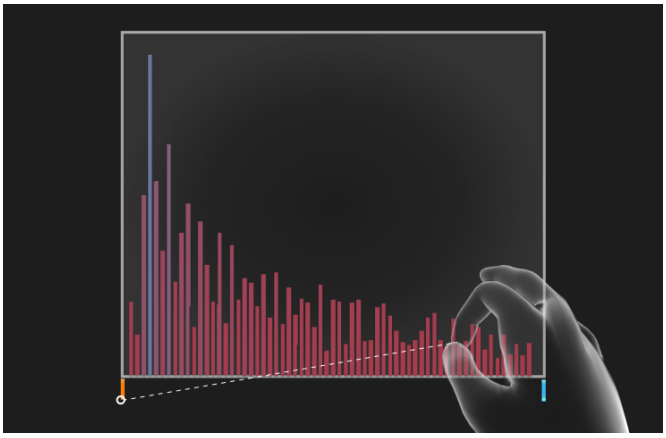
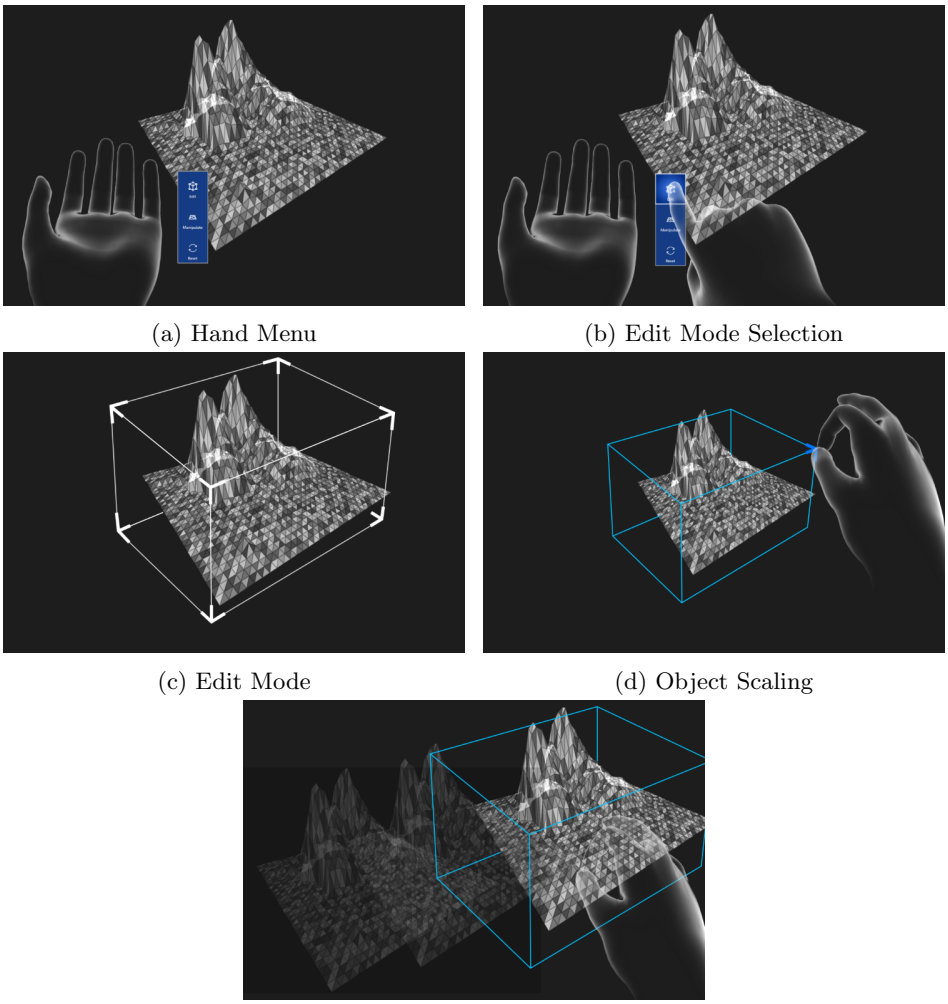


Figure 21: Spectrum Analyzer, Mock-up created in Figma by the author

5.3.7 Object Manipulation

Distinct from Terrain Manipulation, Object Manipulation allows performers the ability to move, scale, and rotate entire terrains (see Figure 22)



(a) Hand Menu

(b) Edit Mode Selection

(c) Edit Mode

(d) Object Scaling

(e) Object Translation

Figure 22: Object Manipulation, Mock-up created in Figma by the author

5.4 Implementation

In this section, we review some of the technical challenges and highlights encountered during development.

In order to implement sound synthesis, the native Unity audio engine was used. The Unity method `OnAudioFilterRead` inserts a custom filter into the audio DSP chain and is called when the system needs to process sample data, which usually happens every 20ms depending on the sample rate and platform. The au-

dio data is an array of floats ranging from [-1.0f;1.0f] and contains audio from the previous filter in the chain or the `AudioClip` on the `AudioSource`. If this is the first filter in the chain and no clip is attached to the audio source, this filter will serve as the audio source. It's also important to note that `OnAudioFilterRead` is not called on the main thread, nor can it be changed to do so, eliminating the use of many Unity functions. This was of particular interest as the `transform` component is required to retrieve the Z-axis of any trajectory lines, yet access to this component is restricted on the audio thread. Therefore, this data had to be collected periodically in the `FixedUpdate` function (called every fixed frame-rate frame), and saved to a variable accessible to the `AudioController`. Fortunately, both functions are called a similar rate, and any discrepancy between the timing would only affect the perceived traversal rate of a trajectory that is being actively moved. However, calibrating these values as well as the audio buffer size is necessary to minimize latency and jitter. In Unity, the size of the DSP buffer can be set to optimize for latency or performance (Best Latency = 256, Default/Good Latency = 512, Best Performance = 1024).

In order to handle multiple inputs, each input component was responsible for processing the audio[12]. In `waveInput`, for instance, the data representing the waveform was processed and inserted into the audio array. If there was `ampInput` from another trajectory signal, then these same values would be modulated by the height of that waveform.

```
void OnAudioFilterRead(float[] data, int channels)
{
    if (waveInput) waveInput.ProcessAudio(data, channels)
    if (ampInput) ampInput.ProcessAudio(data, channels)
}
```

We can see how this was implemented for amplitude processing below. The terrain heights are represented as an array of floats from [-0f;1.0f], so no data transformation is needed to modify the amplitude. However, for the waveform, each height needs to be normalized to [-1.0f;1.0f]. Then, the values are applied across each channel.

```
public void ProcessAudio(float[] data, int channels) {
    bool rawSignal = connectedTo.First().RawSignal;
    var heights = trajectory.GetHeights(upsampling);
    for (int i = 0; i < data.Length; i += channels) {
        if (rawSignal) {
            data[i] *= GetNextRawHeight(rawSize);
            scope.Draw(count % heights.Length, data[i]);
        } else {
            data[i] *= GetNextHeight(heights);
```

[12]Inspired by work from `https://github.com/digego/DisunityST`

```
            scope.Draw(count, data[i]);
        }
        for (int j = 1; j < channels; j++) {
            data[i + j] = data[i]; // mono
        }
    }
}
```

These values are also sent to the scope for additional visualization, as seen in Figure 23

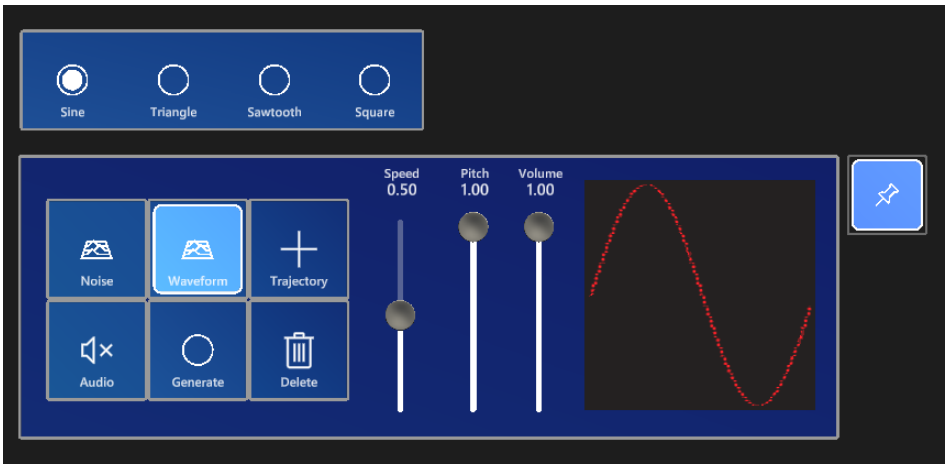


Figure 23: Terrain Menu, captured in Unity

`Pitch` refers to the `AudioSource` parameter, which changes the linear playback speed. Unlike waveform and amplitude, modifying the pitch parameter is only allowed in the main thread. To achieve this, a separate coroutine is executed.

```
IEnumerator ProcessPitch() {
    while (true) {
        yield return new WaitForSecondsRealtime(waitTime);
        if (pitchInput.HasConnection) {
            var val = pitchInput.ProcessTrajectory(terrain.Size, count);
            count++;
            count %= terrain.Size;
            audioSource.pitch = val;
        }
    }
}
```

A significant issue was that terrain resolution had to remain relatively low for performance reasons. This resulted in sample lengths that were frequently much shorter than sample rates. Initially, audio was simply looped, but this resulted in extremely high frequencies. In order to allow for lower frequencies, we implemented upsampling through linear interpolation, seen below.

```
public float[,] GetUpsampledHeights(int upsampling) {
    var heights = WaveTerrain.Heights;
    var biggerSize = Size + (upsampling * (Size - 1));
    float[,] newHeights = new float[Size, biggerSize];
    int c;
    for (int i = 0; i < Size; i++) {
        c = 0;
        for (int j = 0; j < biggerSize - 1; j++) {
            newHeights[i, j] = Mathf.Lerp(
                heights[i, c], heights[i, c + 1],
                (j % upsampling + 1) / (upsampling + 1));
            if (j != 0 && j % (upsampling + 1) == 0) {
                newHeights[i, j] = heights[i, c + 1];
                c++;
            }
        }
        newHeights[i, biggerSize - 1] = heights[i, Size - 1];
    }
    return newHeights;
}
```

The Unity terrain system was initially used to create wave terrains as it provides useful methods for mesh generation and collision. However, a significant shortcoming of the system was the inability to rotate and scale the terrain at run-time. Because of the limited FoV, the minimum size of the terrain was also too large when displayed on the HoloLens. We decided that terrain rotation and scaling were too important, so created a custom terrain system. This presented several challenges: the need for Quaternions to correctly display data on scaled and rotated terrains and performance optimizations of the taxing `MeshCollider` system. Though creating the terrain itself was not too difficult. Below you can see the code used to generate a terrain from a 2D array of heights.

```
private void CreateTerrainFromHeights(float[,] heights) {
    var vSize = Mathf.FloorToInt(TerrainSize);
    Verticies = new Vector3[(int)Mathf.Pow(vSize + 1, 2)];
    var tSizeVector = new Vector3(vSize / 2f, 1, vSize / 2f);

    for (int i = 0, x = 0; x <= vSize; x++) {
        for (int z = 0; z <= vSize; z++) {
```

```
            float y = heights[z, x] * prop.Size;
            Verticies[i] = new Vector3(x, y, z) - tSizeVector;
            i++;
        }
    }

    int vert = 0, tris = 0;
    Triangles = new int[vSize * vSize * 6];
    for (int z = 0; z < vSize; z++) {
        for (int x = 0; x < vSize; x++) {
            Triangles[tris + 0] = vert + 0;
            Triangles[tris + 1] = vert + vSize + 1;
            Triangles[tris + 2] = vert + 1;
            Triangles[tris + 3] = vert + 1;
            Triangles[tris + 4] = vert + vSize + 1;
            Triangles[tris + 5] = vert + vSize + 2;
            vert++;
            tris += 6;
        }
        vert++;
    }

    Array.Reverse(Triangles);
}
```

Another development highlight involves custom behaviors around object manipulation. By leveraging MRTK's `TransformConstraint` class that is associated with the `ObjectManipulator` component, we are able to have full control over object movement. This was used in several parts of the system, such as moving the filter markers, trajectory lines, and panning/scaling a terrain through its noise function. The code snippet below shows how the latter was achieved. In the first method, after calculating the difference between the users current hand position and hand position when manipulation started (triggered by the pinch gesture), we send those values to anything listening to the `onPanned` event, which in this case is a function that updates the noise offset. In the second method, a similar approach is applied, except in this case it is the difference between the scale vectors.

```
// Panning
public override void ApplyConstraint(ref MixedRealityTransform mrt) {
    if (!manipulating) return;
    var inverseRotation = Quaternion.Inverse(parent.localRotation);
    var diff =  mrt.Position - worldPoseOnManipulationStart.Position
    onPanned.Invoke(inverseRotation * diff);
}
```

```
// Scaling
public override void ApplyConstraint(ref MixedRealityTransform mrt) {
    if (!manipulating) return;
    var diff = mrt.Scale - startScale;
    onScaled.Invoke(diff);
}
```

Lastly, terrain manipulation was another notable development highlight. By default, MRTK only tracks object collisions with the index finger, known as the `PosePointer`. However, for a better manipulation experience, we applied hand tracking to the entire right hand, with spherical colliders added to each finger tip (see below).

```
if (HandJointUtils.TryGetJointPose(TrackedHandJoint.ThumbTip,
                                   Handedness.Right,
                                   out pose)
){
    colliderObject.SetActive(true);
    colliderObject.transform.position = pose.Position;
}
```

This allowed more precision, though was also the most computationally expensive, due to the fact that every manipulation action required the `MeshCollider` to be recalculated.

Figure 24: Hand Menu Terrain Settings, Captured in Unity

5.5 Evaluation

O'Modhrain [83] proposed that a VMI can be evaluated from the perspective of the audience, the performer, the designer, and the manufacturer. In this section, we evaluate our MRMI from the perspective of the designer, audience, and performer.

5.5.1 Designer

Overall, use of the Unity engine had its strengths and limitations. The MRTK framework, as well as the UI samples it provides, made it simple to add interaction functionality. When testing using *Holographic Remoting*, usage of the PC's computational power provided a smooth and seamless experience. However, when run directly on the HoloLens, noticeable drops in frame-rate would sometimes occur during more computationally expensive interactions, such as terrain manipulation. As a result, it was critical to repeatedly test with a build to ensure that any performance drops did not negatively impact the experience, or work on optimizations otherwise.

The main limitation of using Unity in this work was its audio engine. Another option considered was to use OSC[13] to send values to an external system, such as a DAW[14], which would then handle the audio synthesis. The benefit of this would be using only the strengths from each system—the ability to create and test compelling interactions in Unity, while leveraging specialized audio systems. However, this would almost certainly lead to increased latency and create dependencies on external systems. To this end, it may be worthwhile to examine the use of WebAudio and WebXR. Existing research has already demonstrated the potential for audio synthesis in the browser[15], and WebXR is quickly advancing. The HoloLens is already capable of rendering these experiences.

While code written to work with MRTK is bespoke, there are several facets of this work that could be used in other systems. For instance, both our custom terrain system and audio synthesizer could easily be applied to applications in XR or otherwise.

5.5.2 Audience

Audience evaluation was accomplished through the response of an improvised performance. Specifically, the performance was recorded, using real-time mixed reality capture through a first-person view, and uploaded online[16]. Participants were instructed to watch the performance and then complete an anonymous sur-

[13]Open Sound Control (OSC) is a networking protocol for communication among computers, sound synthesizers, and other multimedia devices that is optimized for modern networking technology.

[14]Digital Audio Workstation (DAW) used for recording, editing, and producing audio files.

[15]An example of livecoding in the browser: `https://gibber.cc/`

[16]Video available at: `https://youtu.be/IHDZX6-ukL8`

vey. The survey was composed of statements to be evaluated in terms of agreement level with numbers from 1 (strongly agree) to 5 (strongly disagree), adapted from the evaluation criteria proposed by [84]. There was also space for open-ended feedback, where we extract recurring themes or unique comments for further discussion. The full set of responses can be found in Appendix A.

1. *Cause comprehension*: "I felt that the musician's interaction with the instrument was understandable"
2. *Effect comprehension*: "I felt that the instrument provided enough audiovisual information to understand what was happening"
3. *Mapping comprehension*: "I felt there was a clear relationship between the musician's actions and the resulting sound"
4. *Intention comprehension*: "I felt that the musician was able to express themselves well using the instrument"
5. *Error comprehension*: "I felt that the musician's errors (if any) were noticeable"
6. "I was more engaged in the musical expression than usual"
7. "I had a better understanding of the musical expression than usual"
8. "I would be interested in trying out the instrument myself"

In total, there were 29 survey respondents. **Mapping** comprehension scored the highest, with 76% and 24% of participants who reported 'strongly agree' and 'somewhat agree', respectively. This was followed by **cause** and **effect**, which both scored similarly, with the presence of at least one 'somewhat disagree'. The response for **intention** was much more varied, although still positive overall. We speculate if the variation was due to the first-person view preventing an audience-performer relationship, and the occlusion of the performer. This would certainly make it difficult to read intention. **Error** scored the lowest, with the majority of respondents reporting errors as not noticeable. Indeed, there were no major errors in the performance, aside from occasional mis-clicks. Though the nature of the piece as improvisational does not lend itself to error analysis.

Of the three participants that said 'somewhat disagree' for engagement, two of them put 'somewhat disagree' for understanding. While one participant put 'somewhat disagree' for understanding, they put 'somewhat agree' for engagement. It would have been interesting to investigate the relationship between understanding and engagement, though we are unable to draw any conclusions due to the small sample size. In hindsight, it seems the use of engagement may have been limiting. Replacing the statement with "I enjoyed the performance" may have been better. It is also worthwhile to note that all but one participant responded that they would be interested in trying out the instrument themselves.

Overall, the lack of validation for each area makes it difficult to evaluate the instrument in terms of criteria. This could have been achieved through open-ended prompts, asking participants to describe the system, or provide an example of an error, such as was done in [84]. We will certainly take this into consideration for future work.

Table 2: Survey Results (n=29)

Area	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
Cause	45%	45%	7%	3%	0%
Effect	52%	38%	3%	7%	0%
Mapping	76%	24%	0%	0%	0%
Intention	34%	41%	21%	3%	0%
Error	3%	21%	34%	21%	21%

Next, we look at the open feedback section, of which 18 participants responded. The comments were overwhelmingly positive, many noting what they enjoyed, points of constructive criticism, and suggestions for future additions.

We begin with the praise. Some respondents expressed interest in watching the performance live. Sentiment included "cool", "iron man", "beautiful", "sickest performance", "sounds incredible", "brilliant". The criticisms involved the small field-of-view, small text size, audio choppiness, overwhelming visual clusters, and general lack of understanding, with recurring mention about uncertainty on how the "beams of light" affect the music. The majority of the feedback concerned suggestions for future work, compiled below in Table 3.

Table 3: Survey Respondent Suggestions

Another point of view
Ability to add baseline audio track
Presets of specific instruments, melodies, or beats
Perform in different physical environments
Develop for consumer VR/AR devices
Simple UI changes improve understandability
Option to pause music during setup
More spacing of terrains and panels
Minimize parts not currently using to avoid distractions
Prefer to not have UI visible to listener/audience
Reverb and distortion should be more granularly manipulated
Simplifying the instrument in the future for the sake of cohesion and consistency
Graphical scores embedded into the experience and more abstract ways of exciting the instrument

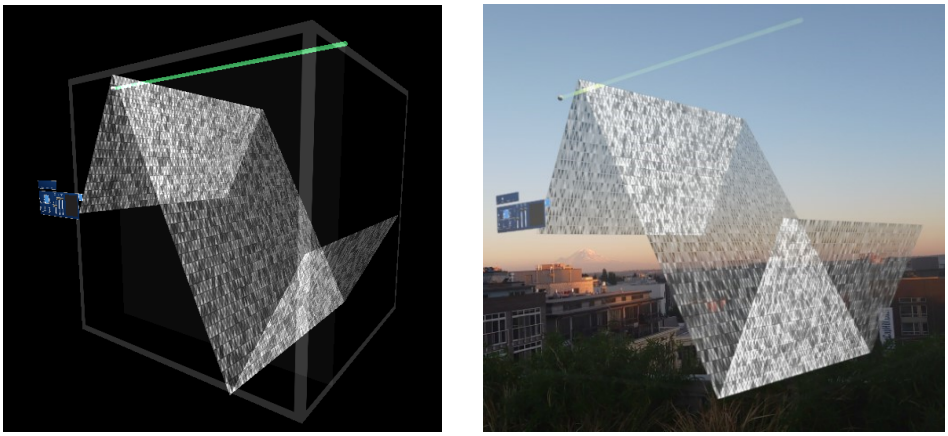
5.5.3 Performer

One major consideration is the effect of Mixed Reality Capture on computational performance. The overhead made the system considerably slower, lending to occasional mis-clicks. There were also audio jitters that were only audible in the video recording. This definitely hindered the performance, and will need to be addressed in the future. The original plan was to use Spectator View to get a third-person view of the performance. However, this would have necessitated the purchase of a second HoloLens or the creation of an Android/iPhone version, neither of which could be completed in time. There were also some discrepancies in visibility due to the light differences between the test and real environment. Having done development in a dimmer room, the presence of sunlight meant that certain materials were barely visible. This was especially noticeable for the terrain borders and trajectory planes, which were intended to aid comprehension. As seen in Figure 25, they are practically invisible when outside.

Despite these limitations, performing with the instrument was both stimulating and enjoyable. Because of the system's generative nature, it was simple to build on previous work. For instance, we can start with a "bass" track by using AM synthesis with a sawtooth wave as the modulating signal, then adding pitch randomization for the "melody", and low-pass filter cutoff modulation for sound variety. We also felt there was a good amount of control, especially when combined with an understanding of basic audio synthesis techniques. Of course, using noise algorithms added a degree of randomization, however these could be utilized in controlled ways. For instance, knowing the difference between Voronoi and Perlin noise, and being able to scale such noise, we able to both predict and plan

the nature of the audio output, yet still benefit from randomization, a controlled chaos.

As mentioned in the previous section, there were no major errors during the performance. At one point, when mapping trajectories, the "Pitch" port was selected, instead of the "Amp" port, which naturally created a totally unexpected sound result. While this action could have easily been undone, it ended up working out, and ultimately benefited the performance. We can attribute this to the nature of the pieces, as experimental works with gradual sound changes. Perhaps such unintended interactions should be welcomed, as they provide an opportunity for the performer to adapt and explore the music in totally new ways.



(a) Editor (b) Outside

Figure 25: Terrain Visibility Comparison

5.5.4 Dimensional Analysis

Next, discussion is structured around the three dimensions from our proposed MRMI framework, supported by reflections from the previous two sections.

Embodiment was explored through the terrains themselves, as interactable audiovisual objects. The potential for multi-process control (with visual feedback) is addressed through audio parameter mapping, allowing one terrain to modulate a virtually infinite number of parameters. Through direct hand manipulation, the system enables users to create and explore musical landscapes in a way that is both intuitive and expressive, offering consistent audiovisual feedback. **Magicality** takes multiple directions in this work. The primary interaction metaphor is wave terrain synthesis. Depending on the user's familiarity with WTS, there will be different perceptions of magicality. However, certainly the interaction of using one's hand to manipulate a virtual terrain is magical in the sense of producing audio, but familiar to a sandbox (natural). Several classic synthesis techniques are

employed, such as amplitude modulation and additive synthesis. Once again, magicality here is dependent on familiarity with these techniques. For someone who is very knowledgeable, the system provides full transparency, as parameters are manually mapped on a one-to-one basis. **Relationships** was explored in several ways. First, the relationships between the terrains were facilitated through parameter mapping. Though, the relationship between the audience and performer was certainly a miss, and with only a sole performer, multi-user collaboration was not considered. There was, however, an innate regard for space. Spatial awareness was enabled on the system, meaning that meshes could "rest" on surfaces, or be visually occluded in 3D space. Furthermore, basic audio localization fostered the relationship between the user's distance from a terrain and it's perceived volume. While the terrains could be expanded and contracted, movement was generally required due to the size of the terrains and the HoloLens' limited field-of-view.

5.6 Future Work

This work demonstrates the potential of MR as a medium for musical expression, and is undoubtedly only the beginning.

As discussed previously (5.5.2), the lack of a spectator view was one of the most significant limiting factors in terms of the performance. Correspondingly, this would be one of the first things addressed, either through the development of a mobile app or the acquisition of a secondary HoloLens for performance capture. On the topic of multiple HoloLens', multiplayer support would definitely bring the system to the next level, by means of physical co-location or virtual avatars. Multi-user collaboration is one of the more powerful applications of MR technologies, and certainly one to investigate in the future. Though this would require a complete overhaul of the system, along with intensive performance optimizations.

In order to allow the performer to start making music quickly, a selection of presets could be offered. For instance, an AM synthesis preset would instantly create a two-terrain system, with one terrain modulating the amplitude of the other. To this end, a grouping feature that allows users to morph between various sub-systems would also be visually compelling. For more musical control, quantization could be explored through the use of visual filters. For instance, dragging a "quantizer" onto a terrain that would bidirectionally affect its sonic and visual representation.

There were a few features that were not fully realized in time for the performance, such as hand-drawn trajectories (see Figure 26). This could be completed and expanded to include circular trajectories, as well as trajectories generated from waveforms.

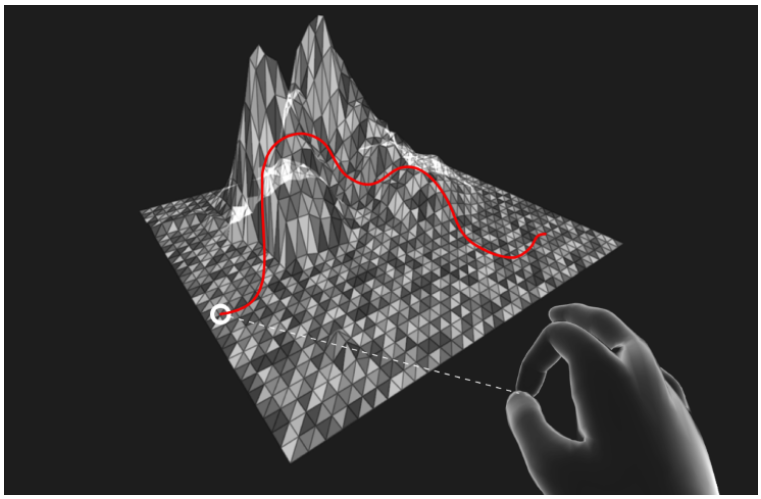


Figure 26: Trajectory Drawing, Mock-up created in Figma by the author

Another was musical gestures. The system currently only allows for gesture mapping at compile-time, which for the performance had been set to "grab" controlling the global volume. However, this could be improved to allow the user to map actions to any sonic parameter, such as the the level of distortion, or the cutoff frequency of a low-pass filter at run-time. A great example of this kind of interaction can be found in [85], which uses interactive machine learning for user-adaptive hand pose recognition. On a similar thread, live terrain height sampling was another feature that didn't make it. It would have allowed the performer to manually adjust certain audio parameters based on the terrain's height at the point of hand contact. For instance, waving through a sine terrain to modify the pitch.

Spatialization also falls under this category. In our system, logarithmic rolloff was used to control the velocity of sound attenuation based on the user's distance from a particular terrain. This could be enhanced further with more localized directional sound, allowing the performer to both control stereo panning with their head and better determine which terrain is producing what sound. Another type of spatialization, usage of the environment spatial mesh as the basis for new terrains remained exploratory, but should definitely be elaborated on in the future.

Evidently, there is no shortage of work to be done. However, with all of these features, adherence to the theme of wave terrain synthesis is crucial, and should guide any future development of this instrument.

6 Conclusion

In the first half of this paper, we explored the meaning of Mixed Reality, its place in the RV continuum, and the other conceptual frameworks that seek to define it. We offer a working definition of a Mixed Reality Musical Instrument (MRMI) as an embodied system for expressive musical performance, characterized by relationships between the performer, the virtual, and the physical environment. Through careful literature review and consideration of the affordances in MR, we distinguish this new class of instrumentation from existing VRMIs. We propose a framework based on three interconnected dimensions to aid NIME practitioners in the design and analysis of MRMIs via a common shared vocabulary. We offer examples of how this framework can be used through the discussion of three different MRMIs. Expert evaluation of the framework found it valuable as a foundational piece for MRMI discussions, while taking issue with the technocentric dialog on VR, and recommending more consideration of musical practice throughout.

In the second half, we detailed the development of a novel MRMI, *Wavelength*, that uses the metaphor of Wave Terrain Synthesis. We evaluate it through audience response of an improvised performance, personal reflection, and based on the dimensions of our proposed framework. The performance was well-received by the audience, scoring high for cause, effect, and mapping comprehension. Open-ended feedback indicated a strong interest in watching a live performance, with numerous suggestions for future re-designs. From personal reflection as the performer, the instrument lent itself well for improvised performance, namely due to its modular and generative nature, which made it simple to build upon. It has been released as open-source software[17], allowing others to contribute to the project and create derivative works. This includes the custom terrain system that addresses the limitations with terrain transform manipulation found in Unity, as well as the synthesizer that utilizes Unity's audio engine, for example.

This work is only the beginning for MRMIs and we hope that our contributions will aid future researchers and developers alike as we unravel the potential for this exciting field together.

[17] `https://github.com/kitzeller/wavelength`

References

[1] Jaron Lanier. "The sound of one hand". In: *Whole earth review* 79 (1993), pp. 30–4.

[2] Lisa Rebenitsch. "Managing cybersickness in virtual reality". In: *XRDS: Crossroads, The ACM Magazine for Students* 22.1 (2015), pp. 46–51.

[3] Simone Grassini and Karin Laumann. "Are modern head-mounted displays sexist? A systematic review on gender differences in HMD-mediated virtual reality". In: *Frontiers in psychology* 11 (2020), p. 1604.

[4] Karitta Christina Zellerbach and Charlie Roberts. "A Framework for the Design and Analysis of Mixed Reality Musical Instruments". In: *International Conference on New Interfaces for Musical Expression*. PubPub. 2022.

[5] Paul Milgram et al. "Augmented reality: A class of displays on the reality-virtuality continuum". In: *Telemanipulator and telepresence technologies*. Vol. 2351. International Society for Optics and Photonics. 1995, pp. 282–292.

[6] Carolina Cruz-Neira et al. "The CAVE: audio visual experience automatic virtual environment". In: *Communications of the ACM* 35.6 (1992), pp. 64–73.

[7] Ronald T Azuma. "A survey of augmented reality". In: *Presence: teleoperators & virtual environments* 6.4 (1997), pp. 355–385.

[8] Cécile Chevalier and Chris Kiefer. "What Does Augmented Reality Mean as a Medium of Expression for Computational Artists?" In: *Leonardo* 53.3 (May 2020), pp. 263–267. ISSN: 0024-094X. DOI: 10.1162/leon_a_01740. eprint: https://direct.mit.edu/leon/article-pdf/53/3/263/1881974/leon_a_01740.pdf. URL: https://doi.org/10.1162/leon%5C_a%5C_01740.

[9] John Viega et al. "3D magic lenses". In: *Proceedings of the 9th annual ACM symposium on User interface software and technology*. 1996, pp. 51–58.

[10] Conor McGarrigle. "Augmented Resistance: the possibilities for AR and data driven art". In: *Leonardo Electronic Almanac* 19.1 (2013), pp. 106–115.

[11] Jannick P Rolland, Richard L Holloway, and Henry Fuchs. "Comparison of optical and video see-through, head-mounted displays". In: *Telemanipulator and Telepresence Technologies*. Vol. 2351. International Society for Optics and Photonics. 1995, pp. 293–307.

[12] Maximilian Speicher, Brian D Hall, and Michael Nebeling. "What is mixed reality?" In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–15.

[13] Luca Turchet, Rob Hamilton, and Anil Camci. "Music in Extended Realities". In: *IEEE Access* 9 (2021), pp. 15810–15832.

[14] Rebecca Rouse et al. “MRX: an interdisciplinary framework for mixed reality experience design and criticism”. In: *Digital Creativity* 26.3-4 (2015), pp. 175–181.

[15] Mafkereseb Kassahun Bekele and Erik Champion. “Redefining mixed reality: User-reality-virtuality and virtual heritage perspectives”. In: (2019).

[16] Teemu Mäki-Patola et al. “Experiments with virtual reality instruments”. In: *Proceedings of the 2005 conference on New interfaces for musical expression*. 2005, pp. 11–16.

[17] Panopticon. *Sounds of the metaverse: A brief history of virtual reality music instruments and virtual music venues*. June 2021. URL: https://panopticon.am/a-brief-history-of-virtual-reality-music-instruments-and-virtual-music-venues/.

[18] *Dancing with holograms: CWRU stages first-of-its-kind mixed-reality dance performance using Microsoft HoloLens*. Feb. 2021. URL: https://thedaily.case.edu/dancing-holograms-cwru-stages-first-kind-mixed-reality-dance-performance-using-microsoft-hololens/.

[19] Rod Selfridge and Mathieu Barthet. “Augmented live music performance using mixed reality and emotion feedback”. In: *14th International Symposium on Computer Music Multidisciplinary Research*. 2019, p. 210.

[20] Dario Mazzanti et al. “Augmented Stage for Participatory Performances.” In: *International Conference on New Interfaces for Musical Expression*. 2014, pp. 29–34.

[21] Victor Zappi et al. “Design and Evaluation of a Hybrid Reality Performance.” In: *International Conference on New Interfaces for Musical Expression*. Vol. 11. Citeseer. 2011, pp. 355–360.

[22] Rob Hamilton. “Coretet: a 21st Century Virtual Interface for Musical Expression”. In: *14th International Symposium on Computer Music Multidisciplinary Research*. 2019, p. 1010.

[23] David Johnson and George Tzanetakis. “Vrmin: using mixed reality to augment the theremin for musical tutoring.” In: *International Conference on New Interfaces for Musical Expression*. 2017, pp. 151–156.

[24] David Johnson, Daniela Damian, and George Tzanetakis. “Evaluating the effectiveness of mixed reality music instrument learning with the theremin”. In: *Virtual Reality* 24.2 (2020), pp. 303–317.

[25] Dominik Hackl and Christoph Anthes. “HoloKeys-An Augmented Reality Application for Learning the Piano.” In: *Forum media technology*. 2017, pp. 140–144.

[26] Shantanu Das et al. “Music Everywhere–Augmented Reality Piano Improvisation Learning System”. In: *International Conference on New Interfaces for Musical Expression*. 2017, pp. 511–512.

[27] Zacharias Vamvakousis and Rafael Ramirez. "Temporal Control In the EyeHarp Gaze-Controlled Musical Interface." In: *International Conference on New Interfaces for Musical Expression*. 2012.

[28] Nicola Davanzo and Federico Avanzini. "Hands-free accessible digital musical instruments: conceptual framework, challenges, and perspectives". In: *IEEE Access* 8 (2020), pp. 163975–163995.

[29] Graham Wakefield et al. "Collaborative Live-Coding Virtual Worlds with an Immersive Instrument." In: June 2014.

[30] Chad McKinney. "Collaboration and embodiment in networked music interfaces for live performance". PhD thesis. University of Sussex, 2016.

[31] Axel GE Mulder. *Design of virtual three-dimensional instruments for sound control*. Simon Fraser University Canada, 1998.

[32] Stefania Serafin et al. "Virtual reality musical instruments: State of the art, design principles, and future directions". In: *Computer Music Journal* 40.3 (2016), pp. 22–40.

[33] Florent Berthaut, Victor Zappi, and Dario Mazzanti. "Scenography of immersive virtual musical instruments". In: *2014 IEEE VR Workshop: Sonic interaction in virtual environments (SIVE)*. IEEE. 2014, pp. 19–24.

[34] Michael E McCauley and Thomas J Sharkey. "Cybersickness: Perception of self-motion in virtual environments". In: *Presence: Teleoperators & Virtual Environments* 1.3 (1992), pp. 311–318.

[35] Kay M Stanney, Robert S Kennedy, and Julie M Drexler. "Cybersickness is not simulator sickness". In: *Proceedings of the Human Factors and Ergonomics Society annual meeting*. Vol. 41. 2. SAGE Publications Sage CA: Los Angeles, CA. 1997, pp. 1138–1142.

[36] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. "A systematic review of cybersickness". In: *Proceedings of the 2014 conference on interactive entertainment*. 2014, pp. 1–9.

[37] Séamas Weech, Sophie Kenny, and Michael Barnett-Cowan. "Presence and cybersickness in virtual reality are negatively related: a review". In: *Frontiers in psychology* 10 (2019), p. 158.

[38] Andras Kemeny, Jean-Rémy Chardonnet, and Florent Colombet. "Getting rid of cybersickness". In: *Virtual Reality, Augmented Reality, and Simulators* (2020).

[39] Alla Vovk et al. "Simulator sickness in augmented reality training using the Microsoft HoloLens". In: *Proceedings of the 2018 CHI conference on human factors in computing systems*. 2018, pp. 1–9.

[40] Robert S Kennedy et al. "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness". In: *The international journal of aviation psychology* 3.3 (1993), pp. 203–220.

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

[42] Konstantina Kilteni, Ilias Bergstrom, and Mel Slater. “Drumming in immersive virtual reality: the body shapes the way we play”. In: *IEEE transactions on visualization and computer graphics* 19.4 (2013), pp. 597–605.

[43] Allen J Fairchild et al. “A mixed reality telepresence system for collaborative space operation”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 27.4 (2016), pp. 814–827.

[44] Mel Slater. “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009), pp. 3549–3557.

[45] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. “The experience of presence: Factor analytic insights”. In: *Presence: Teleoperators & Virtual Environments* 10.3 (2001), pp. 266–281.

[46] Holger Regenbrecht et al. “Mixed voxel reality: Presence and embodiment in low fidelity, visually coherent, mixed reality environments”. In: *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2017, pp. 90–99.

[47] David Drascic and Paul Milgram. “Perceptual issues in augmented reality”. In: *Stereoscopic displays and virtual reality systems III*. Vol. 2653. International Society for Optics and Photonics. 1996, pp. 123–134.

[48] J Adam Jones et al. “The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception”. In: *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. 2008, pp. 9–14.

[49] Bernard C Kress and William J Cummings. “11-1: Invited paper: Towards the ultimate mixed reality experience: HoloLens display architecture choices”. In: *SID symposium digest of technical papers*. Vol. 48. 1. Wiley Online Library. 2017, pp. 127–131.

[50] Luc Nijs, Micheline Lesaffre, and Marc Leman. “The musical instrument as a natural extension of the musician”. In: *the 5th Conference of Interdisciplinary Musicology*. LAM-Institut jean Le Rond d’Alembert. 2009, pp. 132–133.

[51] Inwook Hwang, Hyungki Son, and Jin Ryong Kim. “AirPiano: Enhancing music playing experience in virtual reality with mid-air haptic feedback”. In: *2017 IEEE World Haptics Conference (WHC)*. IEEE. 2017, pp. 213–218.

[52] Pedro Lopes et al. “Adding force feedback to mixed reality experiences and games using electrical muscle stimulation”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–13.

[53] Luc Nijs et al. “Interacting with the Music Paint Machine: Relating the constructs of flow experience and presence”. In: *Interacting with Computers* 24.4 (2012), pp. 237–250.

[54] Florent Berthaut, Myriam Desainte-Catherine, and Martin Hachet. “Interacting with 3D reactive widgets for musical performance”. In: *Journal of New Music Research* 40.3 (2011), pp. 253–263.

[55] Florent Berthaut. “3D interaction techniques for musical expression”. In: *Journal of New Music Research* 49.1 (2020), pp. 60–72.

[56] Christopher Dobrian and Daniel Koppelman. “The ’E’ in NIME: Musical Expression with New Computer Interfaces”. In: *Proceedings of the 2006 Conference on New Interfaces for Musical Expression*. Paris, France: IRCAM — Centre Pompidou, 2006, pp. 277–282. ISBN: 2844263143.

[57] Victor Zappi and Andrew P McPherson. “Dimensionality and Appropriation in Digital Musical Instrument Design.” In: *International Conference on New Interfaces for Musical Expression*. Vol. 14. Citeseer. 2014, pp. 455–460.

[58] Andy Hunt and Ross Kirk. “Mapping strategies for musical performance”. In: *Trends in gestural control of music* 21.2000 (2000), pp. 231–258.

[59] W Andrew Schloss. “Using contemporary technology in live performance: The dilemma of the performer”. In: *Journal of New Music Research* 32.3 (2003), pp. 239–242.

[60] Sidney Fels, Ashley Gadd, and Axel Mulder. “Mapping transparency through metaphor: towards more expressive musical instruments”. In: *Organised Sound* 7.2 (2002), pp. 109–126.

[61] Tim Murray-Browne et al. “The medium is the message: Composing instruments and performing mappings”. In: *Proceedings of the international conference on new interfaces for musical expression*. Citeseer. 2011, pp. 56–59.

[62] Raul Altosaar, Adam Tindale, and Judith Doyle. “Physically Colliding with Music: Full-body Interactions with an Audio-only Virtual Reality Interface”. In: *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 2019, pp. 553–557.

[63] Ronald Azuma. “11 Location-Based Mixed and Augmented Reality Storytelling”. In: (2015).

[64] Mark Billinghurst and Hirokazu Kato. “Collaborative mixed reality”. In: *Proceedings of the First International Symposium on Mixed Reality*. 1999, pp. 261–284.

[65] Steven Dow et al. “Exploring spatial narratives and mixed reality experiences in Oakland Cemetery”. In: *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. 2005, pp. 51–60.

[66] Valentin Bauer and Tifanie Bouchara. “First Steps Towards Augmented Reality Interactive Electronic Music Production”. In: *2021 IEEE Conference*

on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). IEEE. 2021, pp. 90–93.

[67] D Schmalsteig et al. "An environment for collaboration in augmented reality". In: *Collaborative Virtual Environments (CVE), Extended abstract* (1996).

[68] Ivan Poupyrev et al. "Augmented groove: Collaborative jamming in augmented reality". In: *ACM SIGGRAPH 2000 Conference Abstracts and Applications*. Vol. 17. 7. 2000, p. 77.

[69] Ian Thomas Riley. "Touching Light: a Framework for the Facilitation of Music-Making in Mixed Reality". PhD thesis. West Virginia University, 2021.

[70] Thor Magnusson. "An Epistemic Dimension Space for Musical Devices." In: *International Conference on New Interfaces for Musical Expression*. 2010, pp. 43–46.

[71] Andrew McPherson et al. "The M in NIME: Motivic analysis and the case for a musicology of NIME performances". In: International Conference on New Interfaces for Musical Expression. 2022.

[72] Sergi Jordà. "Digital instruments and players: Part ii-diversity, freedom and control". In: *ICMC*. 2004.

[73] Sean Ong. "Beginning windows mixed reality programming". In: *Berkeley, CA: Apress. doi* 10 (2017), pp. 978–1.

[74] *Mixed reality capture overview - mixed reality*. URL: `https://docs.microsoft.com/en-us/windows/mixed-reality/develop/advanced-concepts/mixed-reality-capture-overview`.

[75] *HoloLens 2 hardware — Microsoft Docs*. URL: `https://docs.microsoft.com/en-us/hololens/hololens2-hardware`.

[76] *Holographic remoting overview - mixed reality*. URL: `https://docs.microsoft.com/en-us/windows/mixed-reality/develop/native/holographic-remoting-overview`.

[77] Stuart James. "Developing a flexible and expressive realtime polyphonic wave terrain synthesis instrument based on a visual and multidimensional methodology". PhD thesis. Feb. 2005.

[78] Sohejl Zabetian. "Extensions to Dynamic Wave Terrain Synthesis for Multidimensional Polyphonic Expression". MA thesis. Aalborg University, Sept. 2018.

[79] T. Holmes. *Electronic and Experimental Music: Technology, Music, and Culture*. Taylor & Francis, 2008. ISBN: 9781135906160. URL: `https://books.google.com/books?id=tDaUAgAAQBAJ`.

[80] Curtis Roads, John Strawn, et al. *The computer music tutorial*. MIT press, 1996.

[81] John M Chowning. “The synthesis of complex audio spectra by means of frequency modulation”. In: *Journal of the audio engineering society* 21.7 (1973), pp. 526–534.

[82] Edmund Lai. *Practical digital signal processing*. Elsevier, 2003.

[83] Sile O’modhrain. “A framework for the evaluation of digital musical instruments”. In: *Computer Music Journal* 35.1 (2011), pp. 28–42.

[84] Jerônimo Barbosa et al. “Considering Audience’s View Towards an Evaluation Methodology for Digital Musical Instruments.” In: *International Conference on New Interfaces for Musical Expression*. 2012.

[85] Max Graf and Mathieu Barthet. “Mixed Reality Musical Interface: Exploring Ergonomics and Adaptive Hand Pose Recognition for Gestural Control”. In: *International Conference on New Interfaces for Musical Expression*. PubPub. 2022.

Appendix A Survey Results

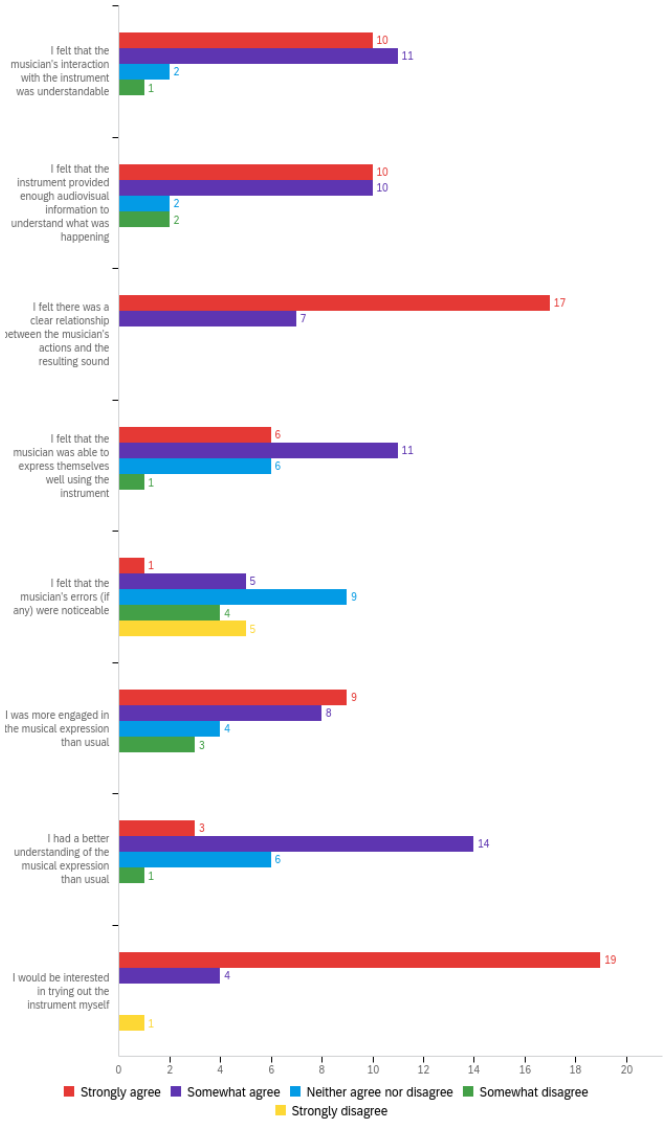


Figure 27: Survey Responses Bar Chart (n=29)

Appendix B System Screenshots

Figure 28: Hand Menu Settings, captured in Unity

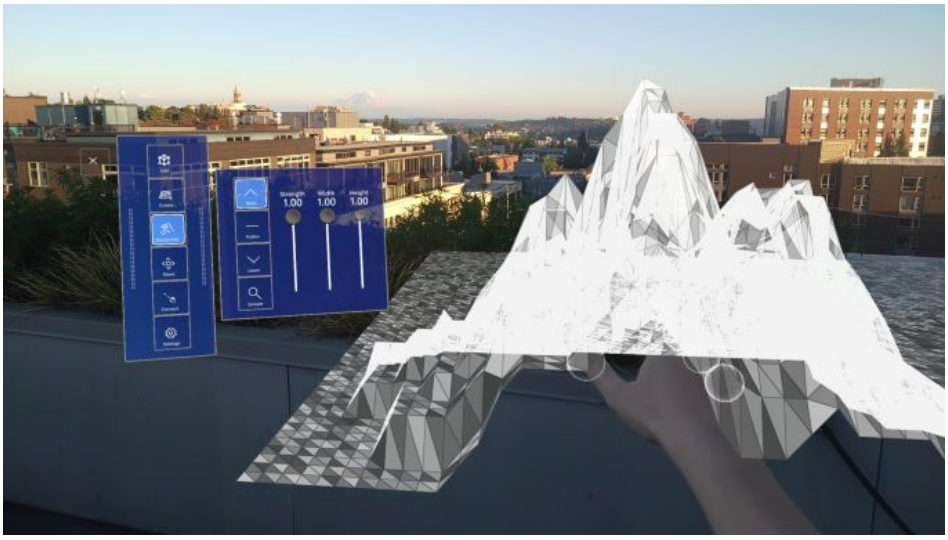


Figure 29: Terrain Manipulation, captured with MRC

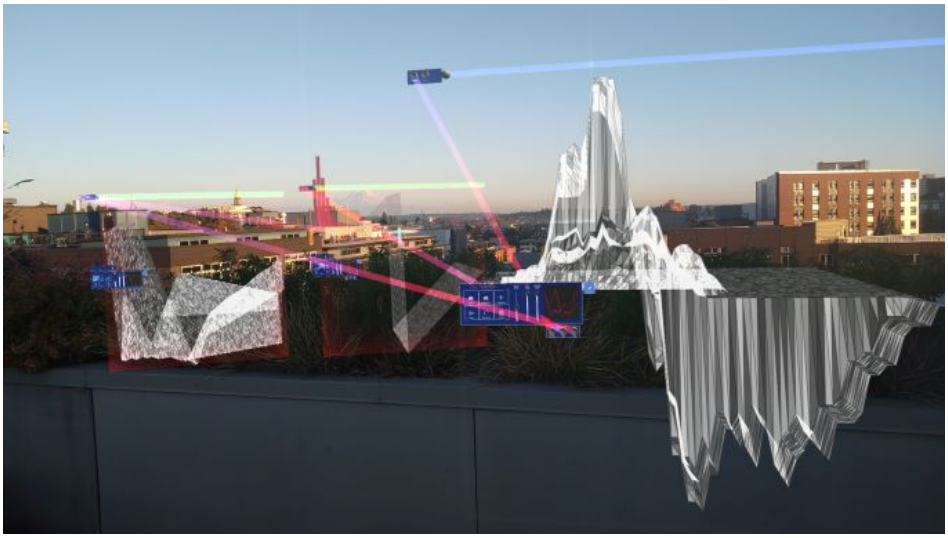


Figure 30: Terrain Modulation, captured with MRC

Appendix C IRB Approval

WORCESTER POLYTECHNIC INSTITUTE

100 INSTITUTE ROAD, WORCESTER MA 01609 USA

Institutional Review Board
FWA #00030698 - HHS #00007374

Notification of IRB Approval

Date:	22-Jul-2022
PI:	Charles Roberts
Protocol Number:	IRB-22-0688
Protocol Title:	The Design and Analysis of Mixed Reality Musical Instruments
Approved Study Personnel:	Zellerbach, Karitta Christina G~Roberts, Charles~
Effective Date:	22-Jul-2022
Exemption Category:	3
Sponsor*:	

The WPI Institutional Review Board (IRB) has reviewed the materials submitted with regard to the above-mentioned protocol.We have determined that this research is exempt from further IRB review under 45 CFR § 46.104 (d). For a detailed description of the categories of exempt research, please refer to the IRB website.

The study is approved indefinitely unless terminated sooner (in writing) by yourself or the WPI IRB. Amendments or changes to the research that might alter this specific approval must be submitted to the WPI IRB for review and may require a full IRB application in order for the research to continue. You are also required to report any adverse events with regard to your study subjects or their data.

Changes to the research which might affect its exempt status must be submitted to the WPI IRB for review and approval before such changes are put into practice. A full IRB application may be required in order for the research to continue.

Please contact the IRB at irb@wpi.edu if you have any questions.