UNIVERSITY
OF TECHNOLOGY
SYDNEY

The Augmented Stage:
A Practice-Based Investigation of Mixed Reality in Live Performance

by Matthew David Hughes

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of
Prof. Andrew Johnston & Prof. Stefan Weinzierl

University of Technology Sydney
Faculty of Engineering and IT

TU Berlin
Audio Communication Group

April 2024

Certificate of Original Authorship

I, Matthew Hughes, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in the Animal Logic Academy at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

I certify that the work in this thesis has not been submitted for a degree nor has it been submitted as part of the requirements for a degree at any other academic institution except as fully acknowledged within the text. This thesis is the result of a Collaborative Doctoral Research Degree program with The Technical University of Berlin.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Matthew Hughes
April, 2024

Acknowledgements

I would like to extend enormous gratitude to my colleagues, collaborators, friends and family who supported me with their minds and hearts throughout the years it took to complete this research project.

A very special thanks goes out to my principal supervisor Andrew Johnston. Without your extended guidance and unyielding encouragement, I would not have completed this document.

To my primary partner of pioneering percussion — Alon Ilsar — I am wholeheartedly grateful. Our works together that make up a significant portion of this thesis have pushed me as an artist and technologist, and have been the source of much pride and enjoyment.

Thank you to my collaborators in the world of physical theatre and dance — David Clarkson and Cloé Fournier. Working together on the novel work presented in this thesis has been fascinating. Your creativity, dedication and knowledge has enriched my understanding of performance and led to some truly impactful outcomes.

To my co-supervisor Andrew Bluff — your work has been an inspiration to me, and your encouragement has fuelled my determination and commitment to this project. To my supervisor at TU Berlin, Stefan Weinzierl, I thank you for your hospitality. The year I spent in Berlin opened my eyes to a wider world of audio-visual art. Your school and colleagues forever changed my attitudes towards computing, software, and freedom.

To my parents Vivienne and John — thank you for your tireless love and support. I have only you to thank for my confidence and ambition. And finally, to my partner in life, Julia — I thank you sincerely for your patience and faith, and for putting up with the stress that accompanied this project. I am lucky to have you by my side.

Abstract

This thesis explores the implementation and impact of mixed reality (MR) technologies within live performance settings. It presents a portfolio of creative productions, each exploring different aspect of MR, including MR presentation through digital projection, the use of augmented reality headsets, multi-modal integration, systems for 3D capture and reconstruction, and systems for networked telepresence and teleimmersion.

The research adopts a practice-based approach through its showcase of seven novel performance artworks that integrate MR technology. The technical and conceptual design decisions made during the development of these artworks are discussed, offering insights into the requirements, considerations, implications and challenges of MR integration in live performance environments. An analysis of documentation including video and photo media, post-performance reflections, field notes and transcribed discussions aids in the presentation of resulting insights, and reveals recurring themes of presence, physicality, connectedness and the blurred boundaries between physical and virtual performance space.

The study culminates with a proposed framework for the development and presentation of MR systems for live performance that incorporates: 1. a set of system design heuristics for MR in live performance, and 2. a suite of modes of interaction to guide the development of performances and systems.

Contributions from this research include seven novel creative productions implementing MR in live performance, a new software for augmented reality telepresence in performance, and a framework guiding practical and conceptual considerations for MR implementation in live performance.

Contents

List of Figures

List of Tables

> Our job is to remind us that there are more contexts than the one that we're in — the one that we think is reality.
>
> — *Alan Kay, pioneer of computing*

1. Introduction

1.1. Setting the stage

Technology relentlessly reshapes our world. With each passing day, the tools we rely on for communication, productivity and entertainment continue to change. This is especially true in the performing arts — a field where novel technologies are frequently adopted and adapted for creative expression. As a technologist and practitioner of live performance, I find myself enthralled by the possibilities that emerging technologies offer in this context.

The development of *mixed reality* has caught the attention of creative professionals worldwide, and is an example of a technology that is poised to reshape the landscape of performing arts. Mixed reality refers to the fusion of physical and virtual worlds. Milgram and Kishino (1994) define it as a spectrum. They call it the *virtuality continuum* — where real environments are placed on one end, totally virtual environments appear on the other, and everything in-between is categorised as mixed reality. This domain has been explored for decades, though the advent of smartphones ushered a new era for mixed reality technology and brought worldwide focus to the concept of *augmented reality* — the real-time superimposition of virtual content onto the physical world (Azuma, 1997).

Augmented reality has made its way into popular culture as the smartphone camera viewfinder proved itself as a boundless digital canvas. Creative *filters* showcase accessible augmented reality for entertainment and communication; mobile games like *Pokemon GO* and *Minecraft Earth* expand the virtual worlds of gaming into physical space; and smartphone augmented reality has provided utility, with brands like *IKEA* allowing customers to view furniture in their homes before purchase.

The pervasiveness of smartphone-based augmented reality does not imply that mixed reality technologies have been examined to exhaustion. On the contrary, the technology and its applications are continuously evolving and developing, and the unexplored possibilities within mixed reality position it as a captivating emergent domain with prospects across diverse fields. With advancements in displays, sensors, networks and processing, mixed reality experiences will continue to become more immersive and impactful.

In the context of live performance, the integration of mixed reality technology holds the potential to redefine the relationship between the performer, the audience, and the performance environment. This technology offers a novel canvas for artistic expression — one that blurs the lines between the physical and the digital. Through mixed reality, performers are able to augment their gestures, create novel forms of audience engagement, extend their craft beyond the physical constraints of traditional venues, and unlock new forms of creative expression. The inclusion of mixed reality technology into the creative palette of live performance artists has the potential to challenge our perceptions of reality, and evoke new experiential and physiological responses from participants. The unexplored transformative potential of mixed reality in the arts is what draws me towards its study, and is the reason for undertaking this work.

1.2. Thesis overview

This thesis explores the integration of mixed reality technologies into live performance. It documents a series of music and dance productions that blend virtual and physical environments to enable performative artistic expression. These productions and their varied system designs showcase a range of interaction modes that mixed reality offers to performers and their audiences.

The core of this exploration has been enriched through collaboration with extraordinary artists. Alon Ilsar has been my partner for the music performances presented, while David Clarkson and Cloé Fournier have provided immense creative contributions with the dance productions. Together we have created some genuinely impactful experiences in an effort to redefine traditional conceptions of performing arts through novel technology.

Over the course of producing the work presented in this thesis, documentation was generated in a range of forms, including videos, photos, notes and transcripts from discussions. This data was captured at various stages: through development, rehearsals, and post-performance reflections. Some semi-structured interviews were conducted to probe further insight, and this collection of documentation guided the understanding of each mixed reality performance implementation. Recurring themes that were drawn out of the data helped to identify and articulate the discoveries that occurred, serving as valuable indicators of the potential, impact, and limitations of mixed reality technology in the field of live performance.

In addition to the creative outputs and their documentation, a new system for performance-focused mixed reality telepresence is presented. The process of designing the software and hardware systems for mixed reality was not predefined, but emerged organically over the course of building each production through an approach incorporating participatory design practices. Several pre-existing software and hardware solutions were employed, tested and sometimes abandoned, marking an iterative process of trial and discovery.

As the criteria for successful mixed reality systems progressively gained definition through my practice, the requirements for *performance* with this technology became increasingly clear. This led to the development of custom software, tuned to meet these particular needs and respond to the challenges encountered as the research evolved. A significant contribution of this thesis therefore, is the current state of this custom-built system for mixed reality telepresence, informed by the practical application and continuous refinement of the research process. The software and its source code is made publicly available[1], and along with its explanation presented in this thesis, this system embodies the insights and experiences gained, offering a tangible resource for future practitioners and researchers in this evolving field.

The aim of this thesis is to explore the integration of mixed reality into the live performance of electronic music and contemporary dance. In order to achieve this aim, the following objectives were identified:

1. Investigate the potential of mixed reality technology in music and dance performances through a series of experimental productions
2. Gain insight through these productions into the challenges, successes and opportunities presented when integrating mixed reality with live performance
3. Develop and refine a performance-focused mixed reality system based on these discoveries

The end results are:

[1] https://git.matth.cc/pointcaster

1. The experimental productions themselves, which form a portfolio of practical case studies
2. A custom-built system for performance-focused mixed reality telepresence, its design considerations, technical details and source code
3. A framework for the design of mixed reality systems for live performance, which includes technical system design heuristics and a set of modes of interaction to guide further development

1.3. Research significance

The outcomes of this research project have the potential to be significant for a variety of stakeholders across disciplines.

The contribution of a rich creative body of work presented in this thesis showcases the artistic possibilities of mixed reality technologies in a tangible and impactful way that will be of interest to the community of artists practising various forms of audio-visual performance. The pieces that constitute the creative portfolio of this research project demonstrate a wide range of mixed reality concepts. A variety of strategies for creating novel technologically-mediated music and dance performances is presented and offers key insights and direction for future artistic ventures in this field.

Similarly, the technical methods used in the design of systems that facilitate these creative productions will also be of interest to audio-visual artists. As mixed reality technologies continue their move towards becoming commonplace, more and more artists will desire to incorporate these contemporary tools into their own work. Performing musicians and dancers, along with artists outside these scenes looking to produce similar immersive experiences, will find the parts of this thesis that detail technical strategies — such as the tools used in producing the artworks — informative and instructional. The work will appeal to computer-based artists, and this thesis stands as a resource that explores and evaluates potential approaches they could apply within their own practice.

This research is also significant to software designers, game developers and researchers of human-computer-interaction that are likewise interested in utilising mixed reality in their own practice. Although the pace of development in mixed reality is steadily increasing, mixed reality as a set of practical technologies for mainstream computing is still in its infancy. The presentation of the software and hardware methods used within this project for the creation of shared mixed reality experiences is significant to professionals and academics interested in mixed reality — but the research additionally explores a long list of related computing topics that includes: real-time graphics manipulation, virtual environment displays, distributed networking practices, software and hardware tools for music composition and performance, 3D scene reconstruction techniques, and interface design.

The discussion and analysis of these technologies, along with the portions of this thesis that document their implementation and elaborate on any shortcomings and practical solutions will be of interest to professionals in these fields. The utilisation of these technologies within the realms of performance art may also have the potential to expose readers from a scientific background to approaches they may not be familiar with, or conclusions they wouldn't have drawn from their own practice. The pointcaster software and its C++ source code is made publicly available alongside this thesis, so that anyone interested in using it, learning from it or contributing to it will have a chance to do so — adding additional value in this thesis' outcomes to those interested in mixed reality software development. It is available at https://git.matth.cc/pointcaster.

1.4. Thesis structure

Chapter 2 (Context) provides a foundational overview to the field of mixed reality. It provides the context necessary for understanding the placement of the work presented in this thesis within the broader research and artistic domain, as well as an introduction to key terms used throughout this document.

Chapter 3 (Methods) explores the methodological approach to this practice-based research project. The design methodologies, development strategies, and data collection and analysis techniques used throughout this research are explored.

Chapters 4 (Augmented Performance with the AirSticks) and 5 (Telepresent Performance) present the creative portfolio of work. These chapters describe the mixed reality productions developed as part of this thesis, and explore their individual themes, system architectures, implementation details and implications on practice. The start of each section presented in these chapters contain links to performance documentation videos. Which should be watched before reading the related discussions on each production.

Chapter 6 (System Design) elaborates on technological choices made during the development of the systems that enabled the creative work discussed in the previous chapters. These system design decisions ultimately led to the construction of a new system for mixed reality performance named *pointcaster*, which is subsequently presented in this chapter. Finally, this chapter also presents a set of *system design heuristics* — the first portion of a framework for the design of systems for mixed reality performance.

Chapter 7 (Modes of Interaction) presents the second portion of this framework: a collection of four modes of interaction that are derived from the mixed reality interactions present in the creative portfolio. This chapter examines the body of creative work through a theoretical lens.

Chapter 8 (Conclusion) provides a concise summary of this thesis.

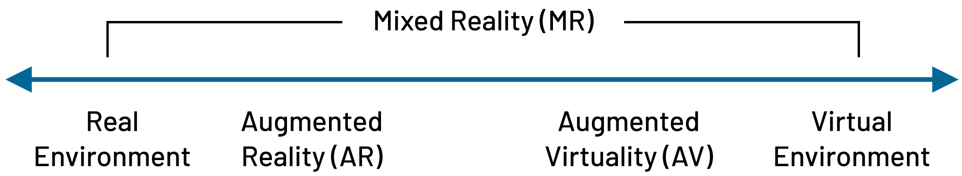
2. Context

As this dissertation investigates the application of mixed reality in live performance, it naturally begins with a review of previous work in this area. This chapter first presents an overview of relevant definitions and research into the broad field of mixed reality, before moving onto its integration within art and performance. It serves as a fundamental primer to ground my own creative portfolio and subsequent discussion within the context of the broader research domain.

2.1. Mixed reality

The concept of *mixed reality* (MR) is defined as the merging of real and virtual worlds. This canonical definition and its introduction in literature originate from the seminal 'A taxonomy of mixed reality visual displays' by Milgram and Kishino (1994). Mixed reality is a general term used to describe computer-aided experiences that can be placed on the *virtuality continuum* — pictured in figure 2.1. This continuum places real environments at one extreme and virtual environments on the other. Any experience that presents a mix of *real-world* (i.e. physical) objects and *virtual-world* objects in the same view falls somewhere in between the extremes of the continuum — in a space referred to as mixed reality. Both *augmented reality* (AR) and *augmented virtuality* are placed within this space on the continuum.

Figure 2.1. – *Milgram and Kishino's virtuality continuum.*



The term *augmented reality* — which Milgram and Kishino describe as the enhancement of a physical environment by computer-generated objects — had appeared in literature prior to its inclusion on the continuum, and contrasts with the implementation of *virtual reality* (VR), where the goal is to immerse the user inside a completely synthetic computer-generated world. Overwhelmingly, the most cited definition of augmented reality is attributed to Azuma (1997), who describes it as a technology that "allows the user to see the real world, with virtual objects superimposed upon or composited with [it]". AR describes computer-aided experiences in which "virtual and real objects [coexist] in the same space" (p. 2), with users maintaining the ability to see the physical world throughout the experience.

Augmented virtuality (AV) defined as the inverse to AR, sits at the opposing end of the virtuality continuum. This term describes experiences that take place in an otherwise virtual environment that is augmented with the presence of components captured from the physical world.

Additionally, the work of Milgram and Kishino provides further depth on MR by establishing a structured taxonomy on the merging of real and virtual worlds. Since the initial introduction of mixed reality as a concept, others have revisited the virtuality continuum, extended Milgram and Kishino's proposed taxonomy, and redefined terms in reaction to the development of modern technologies (Nijholt & Traum, 2005; Skarbez et al., 2021; Williams et al., 2019). While these are all important to note, understanding the definitions of MR, AR, AV and the virtuality continuum will suffice to provide the necessary context for the work presented in this thesis.

For the sake of clarity I will also define the following:

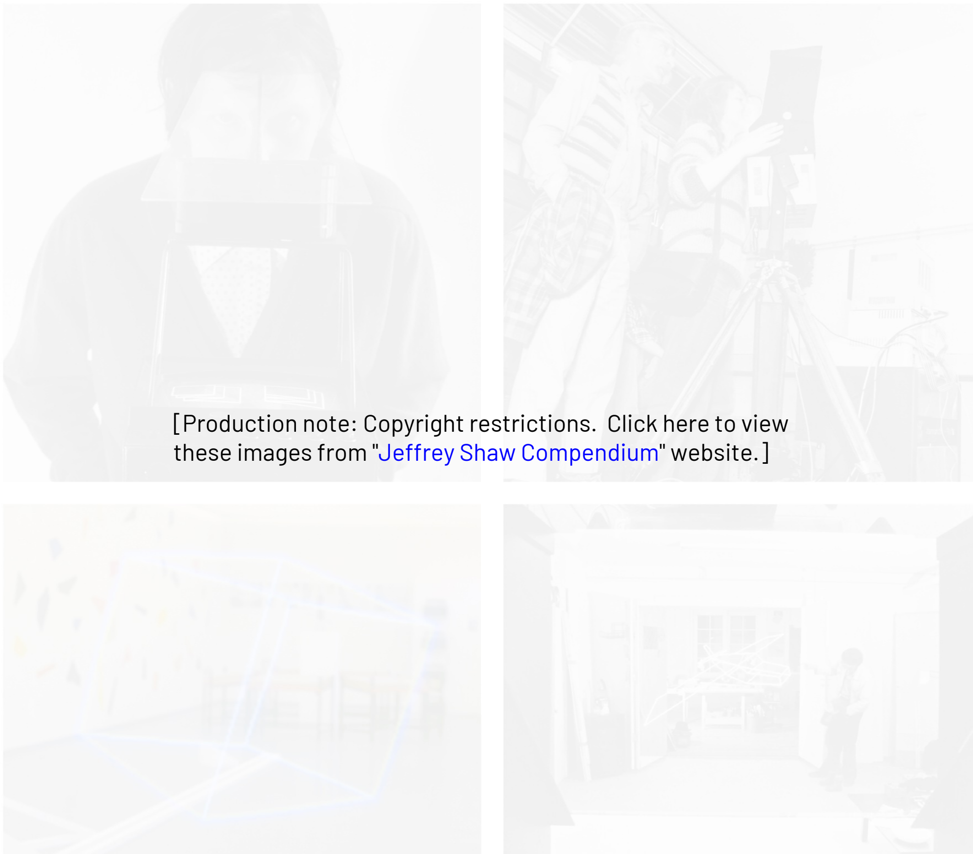
- *Physical environments*, *physical objects* and *physicality* describe entities that exist in our tangible, everyday world. (This is different to much literature that uses the term *real* to describe these entities, which I personally find confusing.)
- *Virtual environments*, *virtual objects* and *virtuality* describe entities that are computer-generated or digitally created.

A *physicality* can be captured and placed into a virtual environment, in which it also becomes a *virtuality* that uses the physical environment as an input. Conversely, a *virtuality* can be introduced into a physical environment through a mixed reality display medium (though it does not simply become a physicality by doing so).

Moving from theory to practice, *implementations* of mixed reality date back to early computing in the late sixties. Sutherland's 1968 presentation of the first head-mounted computer display used a mechanical arm and ultrasonic waves to track the position of the user's head. This data informed the rendered perspective of simple wire-frame images, superimposed onto the user's field of view using miniature cathode ray tubes reflected in prisms in front of each eye to create a stereoscopic augmented reality display.

Some of the first implementations of mixed reality in public artworks can be attributed to the Australian-born Jeffrey Shaw, whose 1970 'Cloud (of daytime sky at night)' projected virtual clouds onto inflatable structures as a precursor to mixed reality *projection mapping* — the practice of projecting virtual images onto

Figure 2.2. – *Jeffrey Shaw's* Virtual Sculpture *in 1981 was one of the first public art exhibits based around an augmented reality interface. Images copyright jeffreyshawcompendium.com.*



irregular-shaped physicalities. His 1975 'Viewpoint', and his 1981 'Virtual Sculpture' (pictured in figure 2.2) fuse the physical museum space with virtual elements to create some of the first augmented reality artworks to be publicly exhibited.

Over the next decades until today, the mediums of mixed reality have continued to develop in various ways across diverse fields, from entertainment and gaming to education and healthcare. The 2010s serve as an inflection point demarcating a significantly expanded reach and understanding of MR technologies.

The proliferation of smartphones, with their combination of increasingly-powerful mobile processors, inbuilt motion sensors, and the ability to superimpose virtual images onto the device's camera feed has placed powerful and portable augmented reality platforms into the hands of billions. Immensely popular mobile games like Pokémon Go, along with camera-based social entertainment platforms such as Snapchat and TikTok, have transformed augmented reality into a technology that is now part of many people's daily lives.

The miniaturisation of personal computing technologies has simultaneously birthed a market of headset devices that offer hands-free mixed reality experiences to consumers. Many virtual reality headsets like the popular Meta Oculus line[1] or Apple's upcoming Vision Pro[2] offer some form of MR experience by passing live video into the device's internal displays. Microsoft has had relative success in enterprise and military markets with their HoloLens[3] augmented reality headset, and several Chinese manufacturers like XReal[4] and Rokid[5] have been capitalising on the availability of cheap micro-displays and optics to create smartphone-powered AR for consumers.

Like AR, augmented virtualities have also become easier to create over the last decade through the democratisation of enabling technologies — in particular, the widespread availability of depth sensors like the Microsoft Kinect[6], the Intel RealSense platform[7], and Apple's LiDAR-enabled devices[8] allow the physical

1 https://oculus.com

2 https://apple.com/apple-vision-pro

3 https://web.archive.org/web/20220517230603/microsoft.com/en-au/hololens

4 https://xreal.com

5 https://web.archive.org/web/20230613084449/rokid.ai

6 https://web.archive.org/web/20230130120318/learn.microsoft.com/en-us/windows/apps/design/devices/kinect-for-windows

7 https://web.archive.org/web/20230701121318/intelrealsense.com

8 https://web.archive.org/web/20230602192317/theverge.com/2020/3/18/21185959/ipad-pro-lidar-scaner-augmented-reality-demo-hot-lava-game

environment to be trivially captured and integrated into virtual worlds. In terms of software, the accessibility and prevalence of game engines and design tools such as Unity[9], Unreal[10] and the open-source Godot Engine[11] have further enabled the creation of mixed reality experiences. These platforms facilitate the rapid development of 3D virtual environments and make the process accessible for practitioners of various technical backgrounds.

2.2. Mixed reality on stage

These recent MR advancements have gifted artists with a new canvas for exploration. The widespread accessibility of these technologies has facilitated the integration of MR into *live performance* across a spectrum of production levels — from research and experimental performance to large-scale, mainstream productions. As a digital artist engaged in live performance, I am particularly drawn to how these technologies unlock novel modes of interaction for practitioners.

The application of mixed reality in live performance forms a rich body of work. It belongs to the practice of audio-visual performance, as well as the broader field of human-computer interaction. With many of the performance works presented in this thesis focusing on the integration of sound and image, the vast fields of visual music and synesthesia are relevant, but are out of this dissertation's scope. The works in the following section are relevant audio-visual performances that focus on the integration between the physical stage and computer-generated components to form mixed realities on-stage.

9 https://web.archive.org/web/20210426085347/unity.com

10 https://web.archive.org/web/20220601183815/unrealengine.com/en-US

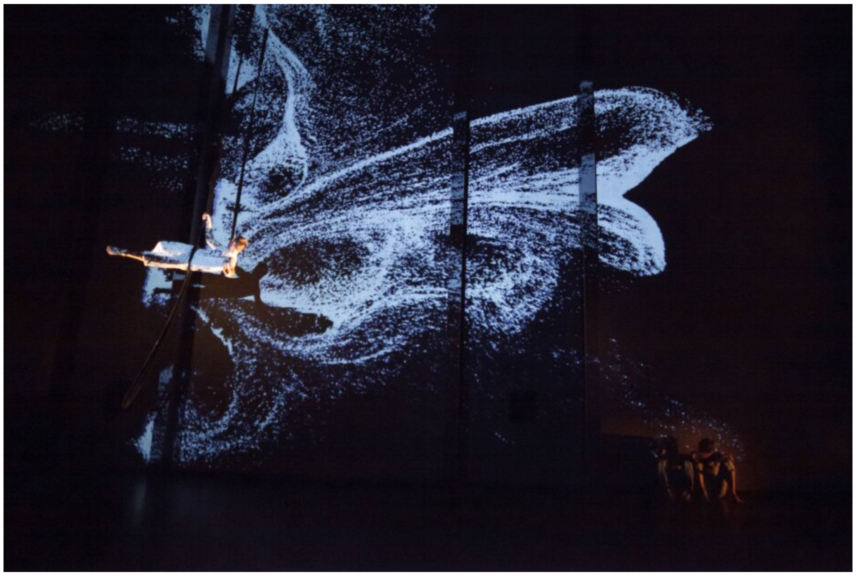
11 https://web.archive.org/web/20230708202435/godotengine.org

In 1992, George Coates' *Invisible Site: A Virtual Sho* pioneered mixed reality performance by projecting computer-generated images onto the stage (Breslauer, 1992). Stereoscopic 3D graphics — such as a giant eyeball that followed performers around — were rendered and presented in real-time inside the performance area using state-of-the-art computers. The technologically-innovative production invited the audience into a 'meta' narrative of on-stage characters experiencing their own virtual reality. This early example demonstrated the potential of presenting a mixed reality among performers on a stage, although interaction shown between the performers and virtual elements was controlled by separate operators off-stage.

A more intimate connection between physicality and virtuality begun to emerge in the years that followed as real-time control systems based on performer movement appeared on stage. The New York-based *Troika Ranch* dance company produced several works through the 90s incorporating devices that used "movements of the performer to generate sound from a MIDI synthesizer, light from computer-controlled devices, or video playback from LaserDisc" (Kepner, 1997, p. 15) to integrate virtual elements and physical performance gestures. The use of movement-based control in these works creates a more direct link between performers and virtual elements, reducing the need for off-stage operation.

The interactivity of virtual components on a stage can also be driven by camera-based systems. Sydney-based *Stalker Theatre* uses infrared cameras to track the motion of physical theatre performers, enabling their gestures to manipulate fluid simulations projected onto the performance environment (A. Johnston, 2013). In some iterations of their system, data from the fluid simulation is fed into a granular synthesiser to additionally generate audio as a result of the participant's motion, creating a synthesis of movement, sound and graphics for a multi-modal mixed reality performance (pictured in figure 2.3). Camera-based motion control is a solution that facilitates interaction between performers and virtual elements through accessible technologies without complex or custom tracking hardware.

Figure 2.3. – *Stalker Theatre's* Encoded *(2012) allows performers to interact with a computer-generated fluid simulation through their motion. Photo copyright Matthew Syres.*

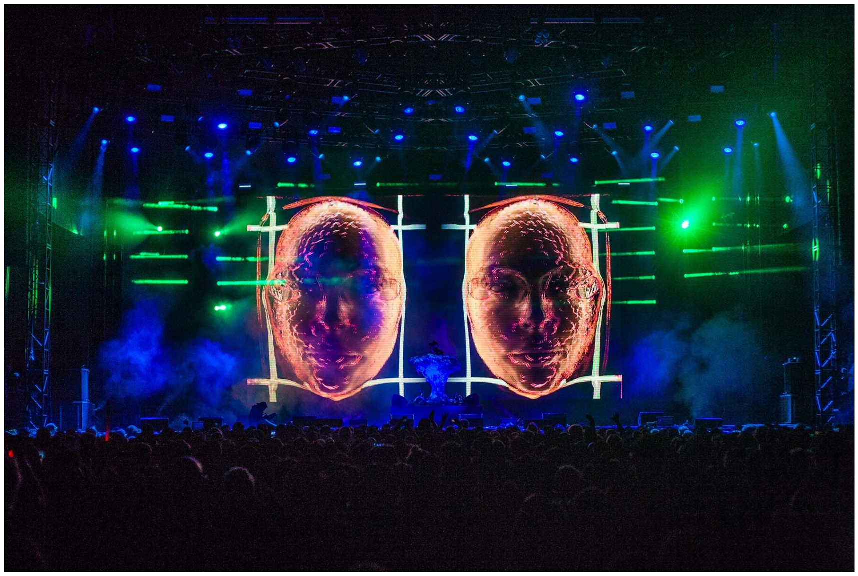


2.3. Mixed reality displays

While motion-based interaction technologies allow performers to control virtual components of a mixed-reality scene dynamically, the impact of these interactions ultimately depends on the technology used to display the visualisations on-stage.

Digital projection of computer-generated images is the most accessible and affordable method of incorporating virtual components into the physical performance space. In contrast, big-budget productions often use LED panels for their superior brightness and contrast — as well as the elimination of any shadows that might appear as a result of front-on projection. Electronic musician Flying Lotus' *in 3D* tour uses stereoscopic LED panels placed behind him to create a mixed reality environment that features a wide array of environments and objects appearing to float beside and around his presence on stage (pictured in figure 2.4). The production features the musician and visual artist improvising together to manipulate the mixed reality environment in real-time (Ip, 2017), and additionally

Figure 2.4. – *'Flying Lotus in 3D' uses stereoscopic LED panels to place virtual elements among the performance area. Photo copyright 3D Live / Flying Lotus.*



incorporates stereoscopic displays — a technology that can be used to increase the 'dimensionality' of virtual elements by presenting separate images to the left and right eyes of the audience.

Another method of enhancing dimensionality is situating virtualities in front of a performer by displaying the images on a transparent surface. The roots of this method can be traced back to traditional stagecraft of the Victorian era. In 1863, John Henry Pepper popularised a method of reflecting off-stage actors onto a pane of glass positioned at the front of a stage to create ghostly illusions that appeared to interact with on-stage performers (England, 2018). This technique — dubbed *Pepper's Ghost* — has been replaced in modern times by direct digital projection onto highly reflective transparent surfaces and fabrics, foregoing the more complex use of mirrors.

In their work *Blue Space*, Walsh et al. (2017) project audio-reactive graphics onto one of these transparent fabrics called *scrim* — which creates a surface for visualisation, situated between the audience and the performer. In this production, Walsh augments her musical gestures by magically blowing particles and strands of coloured light superimposed onto the performance space using her oboe. Musical phrases and their visual representations are additionally morphed through the movement of the oboe's bell.

Alon Ilsar worked with Andrew Bluff in 2015 to produce *Sticks with Viz* — another audio-visual performance that used projections on a scrim to present virtual elements among the stage. Ilsar's electronic percussion (discussed more within upcoming chapters of this thesis) was analysed in real-time by Bluff's software and used as input for a system of physics and fluid simulations to create a kaleidoscopic mixed reality concert (A. J. Bluff, 2017).

Berthaut et al. (2015) used semi-reflective, transparent panes of acrylic as a projection surface placed between the performer and audience to create *Reflets* — their system for mixed reality audio-visual performance. The choice to use a projection surface that combines virtual elements with participant reflections creates a unique display method that is able to situate virtualities within a user's own personal space — as it appears in their reflection. This allows the system to present digital additions inside both the performer's space on stage and the audience space off stage. Combined with cameras used to track participants and allow interaction with the virtual content, this display mechanism facilitates a deeper audience participation.

It is also worth mentioning that in popular culture, these types of transparent projection-based mechanisms for presenting virtual elements on a stage have recently started being described as 'holograms' in order to evoke science-fiction imagery in the general public. Although not technically similar true *holography* — a method involving complex physics to actually project light into 3D space — 'holographic' performances have begun to emerge in music performances and festivals worldwide. The first viral example of an MR performance branded this way was the appearance of a virtual Tupac Shakur at Coachella in 2012. This digital reconstruction was projected onto a transparent polyester sheet to give the appearance of Tupac's presence alongside the real Snoop Dogg on stage (Farivar, 2012). Since this presentation, the ethically-questionable practice of reviving deceased musicians with mixed reality technologies has grown, and it has been possible to book tickets to see Amy Winehouse, Michael Jackson and Buddy Holly among others rise from the dead (Rowell, 2019).

The presentation of images in works discussed thus far have placed virtual imagery onto flat planes at either the front or rear of a stage. However, alternative methods exist for placing images inside 3D space. Jacquet et al. (2007) experimented with placing music visualisation within the audience's proximity in their audio-visual work *3Destruct*. In this installation, projectors shoot audio-reactive visuals onto an array of semi-transparent mosquito net strips attached to the ceiling providing a similar effect to scrim. The strips of netting are divided into four 'volumes' with a projector assigned to each, and a path cut between each volume for the audience to walk through.

Mazzanti et al.'s *Augmented Stage for Participatory Performances* (2014) utilised smartphone-based augmented reality to display visuals that appear around the space of the performers. This work placed image targets behind the stage that were tracked by each audience member's smartphone in order to orient a virtual scene superimposed over their device's live camera feed. Audio parameters of the instruments were mapped to objects in the virtual scene, which were not only reacting to the live music played, but were also interactive for participants who could manipulate the objects through their device's touchscreen.

In section 2.1, I described how AR has become commonplace through the proliferation of smartphones, yet performances like Mazzanti et al.'s are anything but common. There is a clear lack of exploration in the space of smartphone-based AR and live performance. Perhaps this is because distributing a synchronised virtual performance through many audience devices is a complex task, though in 2022 the virtual rock band *The Gorillaz* proved that it was at least possible. While no physical performers were involved in the experience, the virtual characters of the band descended on Times Square in New York and Piccadilly Circus in London to perform their single *Skinny Ape* in AR. Hundreds of people gathered in the city centres to watch the performance together through their smartphone screens as the band members appeared on top of and amongst the buildings surrounding them (Silva, 2022).

2.4. 'Tele-' performance and remote mixed reality

Building upon the concepts introduced in sections 2.1 and 2.2, it is important to the understanding of the creative portfolio presented in this thesis to also provide a brief overview of the MR-adjacent ideas of *teleimmersion*, *telepresence*, and *telematic* performance.

Teleimmersion describes the live embedding of people into a virtual environment (Kurillo & Bajcsy, 2013). Combined with network connectivity, this concept allows for collaborative applications in mixed realities — as digital reconstructions of participants can be placed into *shared* virtual spaces.

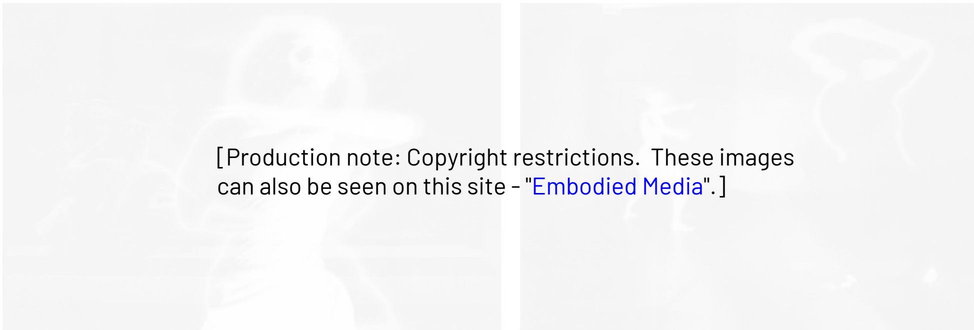
An early example of the implementation of a teleimmersive environment can be found in Myron Krueger's *Videoplace*. Krueger iterated upon this work throughout the 70s and 80s, leading to the exploration of many modes of interaction in such an environment (1985). Participants in *Videoplace* stepped in front of a camera system, which captured their silhouette and placed it into a virtual environment projected in front of them. Once 'inside the computer', the user could use natural human movement and gestures to engage with virtual entities, such as interactive 'critters', a canvas for drawing, and other users that were also placed into the virtual scene from a corresponding setup in the same room, or from an internet-connected remote location.

The union of internet-connected participation with virtual environments enables a concept known as *telepresence*. Telepresence refers to the perception or experience of presence in a location that is physically remote (Draper et al., 1998). Although the definition varies slightly across literature, in this dissertation I use telepresence to describe two related experiences: the sense of being present in a remote location as facilitated by technology, as well as the feeling of a remote participant's presence perceived by those inside a physical space. *Virtuality* in the instance of telepresence, is used as a means of connection and digital transportation — and in the case of performance work, as a means of collaborative artistic expression.

Within the context of performance, the application of these concepts is recognised as *telematic* art. The term *telematic* can include any number of diverse interaction modes that enable geographically-disparate performance through wide area network technology. As telematic works require the live capture of a performance into the virtual realm, followed by its transmission and presentation alongside other virtual or physical entities, all telematic performances inherently exemplify mixed reality.

Telematic performance has a rich history within music and dance. Although remote audio and video communication had been possible for many decades with telephony and satellites, the widespread ability to transmit performance through the internet in the late 1990s spurred greater interest in telematic interaction moving into the new millennium.

Figure 2.5. – *Naugle's* Janus/Ghost Stories *(1999) fused locally captured video feeds with remote dancers sent over a video-conferencing system to create teleimmersive dance performance.*



In 1999, Lisa Marie Naugle presented *Janus/Ghost Stories* — a networked dance performance that used internet video-conferencing technology to connect live dancers between the University of California and Arizona State University (Naugle, 2002). Video projections at each site featured local video merged with remote video to create a viewport for the audience into a teleimmersive dance environment. Pictured in figure 2.5, the video feeds were manipulated in real-time to create dream-like and ghostly interactions.

In 1998, Tanaka and Bongers presented *Global String*, a "musical instrument for hybrid space". A steel cable fifteen metres in length was erected at two remote locations, alongside projected video and audio transmission. The cable extends the virtualities of telepresent musical interaction into the physical realm. As a participant strikes the string in one location, its vibrations are captured and transmitted over the internet to the remote location. This data is used in the disparate site to simultaneously drive a software synthesiser to generate sound, and an electromagnetic actuator that physically excites the string (Tanaka & Bongers, 2002).

Oliveros et al. (2009) presented a series of network-connected musical concerts, one of which contained 44 musicians performing with each other across three sites in the USA. Audio latencies introduced through the transmission make this kind of interaction challenging, though the performers approached the improvisational concert by using the time delays as a feature of the performance itself.

Sheppard et al. (2007) created a 3D teleimmersive system to enable geographically-disparate dance performances in 2006 and 2007. An array made from up to 48 individual cameras was used to capture a dancer's body in 3D, which was sent over the internet between two universities and integrated into a shared virtual space rendered in both locations. The system was used in two public performances that explored the implications of dance choreography within a collaborative virtual 3D space.

While many telematic works approach the virtual space primarily as a faculty for enabling the connection of remote physicalities, interaction with the virtual space itself can be an artistic, exploratory canvas that enables new performative modes of interaction. In this vein, Hamilton (2011) presented multiple musical performances that took place inside complex virtual worlds, and allowed participants to control the production of music itself through interaction with the virtual environment. Performers controlled 'avatars' inside these virtual worlds, using their location in space to advance through musical sequences, shooting virtual projectiles to create auditory effects, and interacting with virtual architecture to trigger and manipulate pre-recorded excerpts of sound.

In the past few years, interest in telematic performance work has increased dramatically. The recent global pandemic placed restrictions around physical collaborative performance for (in some countries) multiple years. While my interest in mixed reality performance was not initially focused on its use-case of telematic art, a realignment was necessary due to the global situation. This is reason for the necessary context provided around the concepts of teleimmersion and telepresence, as they become a strong focus in the second half of my presented creative portfolio.

2.5. Open-air and motion control

A significant amount of the new work presented in this thesis is performed using the *AirSticks* (Ilsar, 2018) — a musical instrument based on handheld motion controllers. This instrument specifically is discussed in section 4.1, though it is worth considering the broader use of *open-air* controllers in live performance. Open-air controllers utilise a range of sensors to afford performers the manipulation of electronic sound through physical gestures. Performers of these instruments control the generation of sound (and potentially graphics, or other software) using the motion and orientation of their hands in the space around them.

The earliest example of an open-air controller system for music is the *Theremin*. Designed by Leon Theremin in 1919, performers of the Theremin move their hands around two antennae, which independently control the pitch and volume of a synthesised sound. The instrument uses changes in electromagnetic fields around the antennae to drive these two parameters.

A range of different approaches to the open-air control of music have been explored since. Of particular resemblance to the AirSticks is the *Buchla Lightning*[12], originally released in 1991. This device uses infrared lights on the tips of two sticks, whose X and Y positions in space are tracked by optical sensors in a tethered box that generates MIDI from this data in real-time, and can additionally output built-in sounds.

[12] https://web.archive.org/web/19980425095643/buchla.com/lightning/descript.html

Other relevant examples of movement-based music control systems include Imogen Heap's *Mi.Mu Gloves* (Mitchell & Heap, 2011), and Atau Tanaka's *BioMuse* (Tanaka, 2000), which both transform gestures of the body into sound. These are part of a lineage of digital musical instruments that are designed around the performer's own idiosyncratic performance practices. In recent years though, it has become easier to build music and performance systems using off-the-shelf handheld controllers not specifically designed for performance. Popular interest in motion control for gaming and virtual reality have led to the wide availability of commodity open-air control systems.

The most prevalent technologies used in the motion controller market today are demonstrated in the controllers for the Meta Quest[13] and Valve Index[14] virtual reality systems. The The Meta Quest controllers emit infrared light similar to the Buchla Lightning. The position of these lights — as determined by cameras on the accompanying headset — is combined with data from an inertial measurement unit (IMU) to calculate orientation and track motion. Conversely, the Valve Index requires infrared light emitters to be placed around the play-area, and the controllers themselves contain sensors that track the projected light to calculate their position and orientation. The accessibility of commodity motion controllers has led to their use within new performance work to enable open-air control of sound and graphics. The *AirSticks* — which are used throughout this thesis' creative portfolio — are themselves based on commodity virtual reality controllers.

Now that an appropriate context for the works presented in this thesis has been established, I will move on to the discussion of methods used in the research.

---

[13] https://web.archive.org/web/20201006100930/forbes.com/sites/joeparlock/2020/10/05/a-look-at-the-oculus-quest-2s-new-touch-controllers

[14] https://web.archive.org/web/20240121091033/liamfoot.com/valve-index-controllers-in-depth-review

3. Methods

3.1. Research through design

Given that this research is concerned with the development of new interactive technological systems, it naturally falls within the domain of Human-Computer Interaction (HCI) research. This interdisciplinary field investigates the design and implementation of computing systems crafted for human use, as well as the study of the phenomena that surround them.

This thesis applies a Research through Design (RtD) methodology, an approach commonly used within the HCI research field. The aim of the research is to push boundaries of live performance through the creation of new works and systems. RtD — as a form of practice-based research — is embedded in this process of creation.

RtD encapsulates the iterative creation and evaluation of *design artefacts* as a means of generating knowledge. In this process, the act of 'making' is regarded as a method of inquiry (Zimmerman et al., 2007). This is particularly relevant in areas of research that involve novel technologies, or the implementation of technologies in novel contexts. Where previously established knowledge may not be directly applicable, new insights can be generated through the act of design.

The design artefacts presented through this thesis encompass a range of outputs. The mixed reality systems presented over the next chapters are design artefacts themselves — as are the creative pieces of work produced using these systems. Artefact contributions of HCI research are not merely end products, but also embody the knowledge and insights gained throughout the design process (Gaver, 2012). The work elaborated on in chapters 4 and 5 embody both practical outcomes and conceptual conclusions. The presentation of each work and their underlying systems form a portfolio of discrete artefact contributions.

In addition, chapter 6 expands the scope of discussion to a wider examination of system design with respect to the findings of the creative portfolio on the whole. A new system for telepresent performance in augmented reality is presented through the context of iteration and evaluation of the entire design journey. This new system's presentation, positioning within the larger research context, and its 'real-world' output (a functioning system and its source code), together form another significant artefact contribution.

Wobbrock and Kientz (2016) define seven types of contributions HCI research can yield, and artefact contributions are only one of them. This thesis also presents outcomes in the form of *theoretical contributions*.

A framework is derived from the iterative design and development process, which consists of two components: a set of system design heuristics, and a suite of modes of interaction. These synthesise practical knowledge gained through the creation of artefacts into a structured criteria, and offer a guidance for designing and implementing mixed reality systems for live performance.

3.2. Practice-based research

RtD aligns naturally with *practice-based research*, which guides the research process of this thesis. This methodology is similarly concerned with the creation of output artefacts, and is often employed by researchers in the intersection of technology and the creative arts (A. J. Bluff, 2017; Ilsar, 2018; A. J. Johnston, 2009).

Practice-based research involves an original investigation with the aim of generating new knowledge through practice and its outcomes (Candy, 2006). When a particular creative outcome contains an original contribution to the field, the product that demonstrates this contribution — be it a composition, an exhibit, a performance, or any number of practical works — is integral in being able to understand the research.

With arts-focused research, the study takes place with the exploration, discovery and change of artistic processes — but in contrast with sole artistic exercise, the artefacts of a practice-based researcher are complemented by documentation that provides context, criticism, analysis and explanation in order to express how the practice itself has contributed to the field. The placement of the artwork within a scholarly context is what "unites the artistic and the academic in an enterprise that impacts on both domains" (Borgdorff, 2012, p. 143).

When combining research with artistic practice, the act of creating new works cannot be detached from the research process. It is not only the manufacture of a novel product, but proof of new insight and understanding framed within the context of practice that defines an artistic practice-based study.

In addition to grounding the research inside real-world application, a practice-based approach is ideal for study that aims to preserve the natural dynamics and spontaneity of artistic creation processes.

> To be practice-based is to be more flexible; more embedded in the immanence of operations; more responsive to material and conceptual change; more reflexive. (Phillips, 2010, p. 71)

In her discussion on different models of creative practice-based research, Candy (2011) outlines four possible outcomes: *works* (i.e. artefacts), *design criteria*, *conceptual frameworks*, and *results* that arise from evaluation. In this light, the already stated HCI outcomes of artefact and theoretical contributions are solidly grounded in practice-based research.

With this in mind, the approach of this thesis is to incorporate components of practice-based research and research through design. This synthesis allows for the dynamics of artistic practice to flourish while examining its processes and output within a research context, contributing outcomes to both the academic and artistic domains.

3.3. Collaborative design

Throughout the creative portfolio of works presented in this thesis, I assumed the role of system designer. If not working alone in this role, as in *Critical Path* (5.4), I was on a team sharing this task, as with the *AirSticks* (4) and *Layer* (5.3). I also took on the roles of software engineer and graphics designer. Only in one case (5.2) do I additionally take on the role of performer, yet the performer is one of the most important users of a system *built for live performance*. In this context, the performance artists are the main stakeholders dependent on the success of the system.

In the example of my work with choreographer and dancer Cloé Fournier for *Critical Path* (5.4), I constructed a system for telepresent dance performance using augmented reality glasses. One approach in this context would be to develop the system, hand it off to the user, and have them compose a performance using it. After which, the system could be evaluated using a measurable usability criteria and potentially iterated upon. This is a totally valid approach in the design and development of an artefact, and encompasses the principles of *user-centred design* — in which the design process engages with the end-user to develop a system with their requirements in mind (Mao et al., 2005).

While an approach that involves first the creation of prototypes before handing off for usability testing does consider the end-user, it potentially overlooks engaging with their unique needs, context, and creative vision from the very beginning. It is a less streamlined route to the same design goal, perhaps missing critical nuances along the way. This is particularly salient in the context of performance art, where the artist's unique style and conceptual ideas drive the creative process. In this context, a more collaborative and interactive design process is beneficial, and the approach of *participatory design* is relevant.

Participatory design is a user-centred design methodology that advocates for the involvement of all stakeholders in the design process to ensure the result meets their needs (Hanington & Martin, 2012). This approach to design emerged as a means of designing human-computer interfaces that engaged workers in the process of designing the systems they would make use of every day (Spinuzzi, 2005), and is centred on the principle of co-creation. It involves key stakeholders in the design process of an artefact to produce systems that are usable, beneficial, and relevant to their contexts.

A major benefit of participatory design is that it engages with the *tacit* knowledge of the stakeholders. The potential richness of understanding that is to be gained from engaging a creative practitioner's experiential knowledge — including insights that may not be easily verbalised — is too valuable to ignore.

> When we think of knowledge, we often think of explicit forms of knowledge: things that are written down, defined, categorized, systematized, or quantified. But to understand knowledge-making in participatory design, we have to understand that much knowledge tends to be tacit. Tacit knowledge is implicit rather than explicit, holistic rather than bounded and systematized; it is what people know without being able to articulate. (Spinuzzi, 2005, p. 165)

This research applies an approach drawing upon the core principles of participatory design, respecting the creative insights of participants to inspire and guide the design process. Returning to the case of my collaboration with the choreographer mentioned above, participatory design respects their existing artistic practice. By incorporating aspects of participatory design, the need to 'work around' any design elements that might contradict their usual practice is minimised. The collaborator is not forced to either simply accept or reject *my* design decisions, but is treated as an active participant in the design of the systems themselves.

Throughout the creative portfolio presented in this thesis, I actively engaged with musicians, dancers, choreographers, and directors in the design process of the interactive systems. Practitioners involved in system design sessions provided input and feedback, while I adapted the technology to their needs and creative vision. The use of real-time visualisation systems and software that can be reprogrammed on the fly is particularly beneficial in this context, as it allows for adaptability and immediate implementation of ideas generated during the co-design sessions. This collaborative approach not only results in more usable and relevant systems, but also fosters a sense of ownership and engagement among the performers, as they have actively contributed to shaping the technology they will use in their creative practice.

Certainly the primary focus of the design approach in this thesis was through collaborating with creative practitioners, though section 4.4 also opens a dialogue with *audiences* during the design process, and begins to explore their importance as stakeholders as well.

3.4. Approaches to development

Each production within this thesis was created in periods of collaborative development, where each member of the team would work concurrently, in the same room or studio or theatre. Even when working on solitary tasks, the shared physical space facilitated an exchange of ideas and creative inputs.

A tangible example of how the participatory approach guided the design is presented in figure 3.1, which shows a *Kanban* board used during the *Critical Path* residency period. Initially developed by Toyota in the 1940s to streamline manufacturing, Kanban revolves around visual management, continual improvement and flexibility. It makes use of cards representing individual tasks within a project that are moved along a board as their status changes. Over the past few decades, Kanban has seen widespread application in software development teams, often in conjunction with *Agile* software development methods (Brechner, 2015).

While we didn't strictly follow the traditional structure of the Kanban method, we utilised its core philosophy to facilitate our participatory process. It allowed for flexible adjustments to tasks and priorities relating to the design of both our system and our performance, while keeping stakeholders informed on design processes. Software development tasks, choreographic tasks and conceptual tasks were combined together on the board. Importantly, the visibility provided by the Kanban board stimulated collaborative discussions and decision-making — ensuring the developed systems aligned with each participant's practice.

Figure 3.1. – *A modified Kanban board was used for visibility between collaborators working on different components of* Critical Path *(5.4). Blue cards represent software development tasks, pink cards represent choreographic tasks, and green/yellow cards represent concepts to explore.*

The approach to design of new systems grounded in practice and active stakeholder participation, as evidenced in our shared physical workspace and use of the Kanban board, is emblematic of *adaptive development*. This is a term taken from the world of software, where all approaches to development range from *predictive* to *adaptive* (Boehm & Turner, 2009).

Predictive development focuses on discretely outlining the system's requirements from the start at both high and low levels, moving forward by developing for these needs and catering to known risks. In contrast, adaptive approaches to development first outline higher-level goals, but rather than sticking to a rigid path of low-level tasks, they are poised to cater to emergent requirements that are not pre-specifiable.

This inherent flexibility makes adaptive development a suitable partner to practice-based research, allowing for low-level system requirements to be discovered and adapted to through the ongoing creative practice and input from collaborators, rather than being rigidly pre-specified. For example, when developing the audio-visual systems for *Computer Storm* with Alon Ilsar (4.4), higher-level motives such as crowd participation and multi-modal integration were specified, along with thematic notions of voyeurism and data autonomy.

Tasks related to the design and implementation of the actual interactive systems emerged as we dove into the creative process in active dialogue with one another. New requirements like bi-directional communication between visualisation and sound systems emerged as we explored creatively. Desires, requirements and implementations of our systems evolved through a process of iterative development, testing and refinement.

3.5. Data collection

As previously stated, practice-based research is not confined to the creation of the artefact alone. The contextualisation, evaluation and reflection surrounding the practical output are equally significant in generating a comprehensive understanding of the research (Candy, 2006). Consequently, a practice-based researcher must generate raw data for evaluation and reflection in ways that properly capture the exploratory nature of their artefact's design process.

For my own data collection I leveraged a range of tools: video recordings to document the performance and development stages of each artefact, comprehensive notes taken throughout the design process, and transcriptions of both semi-structured interviews and unstructured discussions involving participants engaged in the design process.

Video documentation

The creation of audio-visual performance artefacts necessitates comprehensive video documentation. In the context of live performance, capturing video becomes indispensable as it serves as a proxy for the lived experience within the research. Readers of this thesis would likely not have attended most presentations of the works discussed, so the video excerpts presented through chapters 4 and 5 serve as a stand-in. Auslander (2006) eloquently explains the relationship between performance-focused research and its documentation:

> ... performance is always at one level raw material for documentation, the final product through which it will be circulated and with which it will inevitably become defined ... the photography ultimately replaces the reality it documents. (p. 3)

In addition to this, video documentation of performances facilitates the post-performance evaluation and reflection alluded to earlier. For works presented more than once, recordings can chronicle their evolution.

A large collection of video was also captured to thoroughly document the collaborative development processes. 'Behind-the-scenes' clips were frequently recorded, and include moments of the design process, prototype creations and practice sessions — all working to capture the collective journey and evolution of each artefact. As previously mentioned, the development methodology was non-prescriptive, allowing for design tasks to emerge and evolve throughout the process. Capturing video throughout development was essential in supporting a proper understanding of the design journey.

The following is an example of the kind of video recordings that were captured during development.

► **Video 3.1** - *Footage captured from augmented reality headsets during* Critical Path *(5.4) shows experimentation with a prototype system for telepresence.*

Watch at phd.matth.cc/dev-doc-example.

Semi-structured interviews

To generate raw data relating to the works presented in chapter 5, the creative participants of the works *Layer*, *My House Your House* and *Critical Path* took part in post-performance interview sessions.

Interviews can provide the empathetic data collection necessary for analysing practice-based research. Structured interviews — similar to surveys, where an interviewer does not deviate from a pre-determined set of questions — are less flexible than other types of interviews. They often limit the in-depth exploration of topics required when interviewing a small sample group. As such, the more free-form approach of *semi-structured* interviewing was employed, as described by Bryman (2012).

With semi-structured interviewing, the interview session is somewhat planned prior to being carried out through the construction of an *interview guide*. The interviewee is given major flexibility in their response, but the interview guide — a set of specific questions or topics noted for the investigation — provides some direction to the conversation. The open-ended nature of this technique allows the interviewer to dive deep into topics that arise spontaneously during the interview, revealing phenomena that would have otherwise gone undiscovered.

In interviews conducted for this thesis, the semi-structured approach was extended with the application of *video-stimulated recall*. This method involves the presentation of video featuring the interviewee themselves, after (or sometimes during) which they are able to reflect on their own behaviour in the recording (Nguyen et al., 2013). Prepared videos were included in the interview guide in order to focus the discussion on particular moments of practice.

A sample interview is provided in Appendix A, which demonstrates the free-form nature of the semi-structured interview process, as well as the video-stimulated recall method.

Unstructured group discussions

In addition to the interviews, two unstructured group discussions were transcribed. One included a group of ten participants, including myself and the performer after a development showing of *Computer Storm* (4.4). The other was a question-answer session after a showing of *Layer* (5.3), which included the production team of sixteen, an MC, and questions from the audience that just watched the performance.

Group sessions like this — in contrast to the one-on-one approach I have employed for semi-structured interviews — facilitate collective discussion that allows participants to reveal shared or contrasting views. These sessions captured immediate post-performance impressions and communicated the experiences of performers, technologists and audience members. The unstructured nature of this open-ended discussion is considered a *naturalistic* style of ethnographic research, in which the social world is mostly undisturbed as it is studied (Atkinson, 2007). Spontaneity guides the the conversation and shapes the resulting data.

A transcript of one of these group discussions is provided in Appendix B.

Field notes

Lastly, the collection of *field notes* formed an indispensable method of data collection for this research. Notes are a form of auto-ethnographic observation that is fundamental to practice-based research where the researcher is deeply involved as a participant. They serve as the main tool in capturing a practice-based researcher's first-hand experiences, thoughts, insights, and perspectives.

In the context of this research, field notes played a crucial role in documenting the myriad details that video and audio documentation may have missed. These notes were recorded on my smartphone or laptop, and included a mix of descriptions, reflections, and analytical thoughts that arose either in real-time during the creation of the artefacts, or as post-action reflections. They helped create an intricate account of the evolution of each design artefact.

3.6. Thematic analysis

To make sense of these various data sources, an analysis stage was carried out. Analysis methods are fundamentally about data reduction (Bryman, 2012), where the body of gathered information must be narrowed to end with something meaningful and insightful. The approach to this research has drawn from methods present in *thematic analysis*, in which salient concepts present in the data are unearthed through the systematic extraction of *themes* at different scales of inquiry, from overarching global themes, to specific sub-themes and even smaller *codes* (Attride-Stirling, 2001).

Nowell et al. (2017) describe thematic analysis a procedure involving the following steps: familiarisation with the data; generation of codes; finding, reviewing and defining of themes; and the production of a report. It is an example of an inductive approach to research, where findings emerge through the data itself, and research does not begin with pre-defined hypotheses or theories.

In this thesis, when data was collected in the form of semi-structured interviews and recorded discussions — as in 4.4, 5.3 and 5.4 — transcriptions were imported into GNU Emacs[1], where the `orgqda`[2] package was used to organise, navigate and code the data. After coding, notable themes were identified, and further reduced to form the basis of: a) the discussions presented in the relevant portfolio chapters and b) the design framework components discussed in chapters 6 and 7. Each of the system design heuristics and modes of interaction explained in these findings initially emerged from a thematic analysis process.

Coding

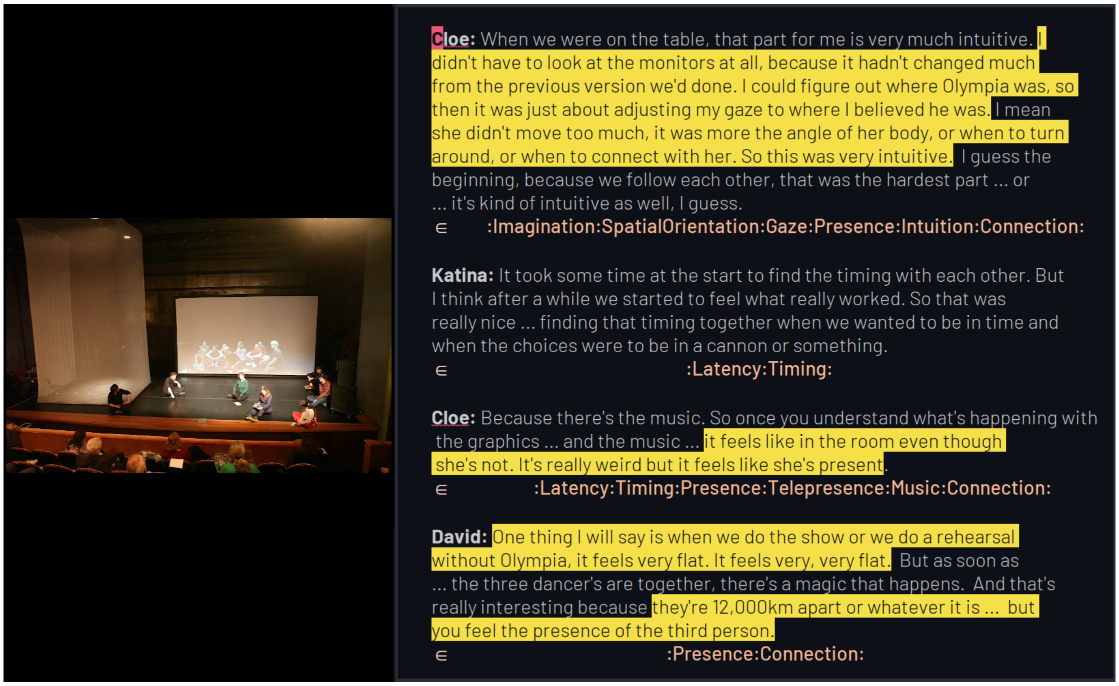
After the initial step of familiarisation with the data — a process that includes repeated reading/viewing of the material, and the possible generation of more notes upon reflection (Nowell et al., 2017) — the initial *coding* procedure takes place. Textual data is analysed line-by-line, and a 'code' is created for any new concern that arises in the text. Codes can start at very low levels, and are initially generated liberally to capture the concepts contained within the data — even for concerns that deviate from the dominant narrative in the analysis (Braun & Clarke, 2006). Coding is a repetitive process however, and passages of text can be assigned and unassigned multiple codes as the analysis progresses. This process transitions the unstructured source material into the beginnings of an understanding of what the data pertains to.

Figure 3.2 shows the initial coding process as carried out using `orgqda`. The source video is presented alongside the audio transcription in an Emacs window, where passages are tagged with codes below their place in the text.

[1] https://www.gnu.org/software/emacs

[2] https://gitlab.com/andersjohansson/orgqda

Figure 3.2. – *A screenshot of the initial coding process using orgqda inside Emacs. The source video is on the left. The transcription on the right contains highlights that were marked up during the familiarisation phase. Codes appear in bold orange text underneath passages they are assigned to.*



Developing themes

After the initial 'code list' is generated, higher-level *themes* are derived. According to DeSantis and Ugarriza (2000), a theme in qualitative data analysis is "an abstract entity that brings meaning and identity to a recurrent experience" (p. 362). Passages of data are recontextualised within the context of their respective codes. Each coded excerpt is organised and subsequently visualised in relation to others sharing the same codes. Higher-level themes are abstracted from lower-level codes, and summaries are written to describe what the themes represent. Once themes are derived and described, they are reviewed in relation to one another and their source. Some collapse into each other, and others are broken down into multiples. Progressively, the themes are refined into a succinct and manageable set that effectively summarises the data (Attride-Stirling, 2001).

Code relations can be inferred from the initial code list in order to recontextualise the data to aid emergence of themes. Figure 3.3 shows `orgqda` visualising data in the context of these code relations. In this figure, we see a view of all the excerpts of an interview that have been tagged with the same two codes — *Connection* and *Presence* — so that we can understand better the relationship between these concepts and tangential codes in the process of deducing themes.

After themes are identified, thematic analysis usually culminates in the production of a report. A thematic report should ideally provide a coherent and engaging narrative that encapsulates the data and themes arising from it (Braun & Clarke, 2006). Verbatim quotes are integrated into the body of the reporting. This is common in the presentation of a thematic analysis, where short quotes and full-length passages are often used to aid understanding of specific points (Nowell et al., 2017).

For this thesis, the identified themes are discussed throughout the creative portfolio, and the data is encapsulated in the body where it is available. The result of this structured approach to understanding the qualitative data was used in the process of synthesising the findings presented in chapters 6 and 7.

3.7. Summary

In summary, this research project employs a *practice-based* approach, characterised by an iterative process of creating, evaluating, and reflecting upon a series of mixed reality performance works. This methodology — situated within the intersection of art, technology and academia — falls under the umbrella of *Research through Design*.

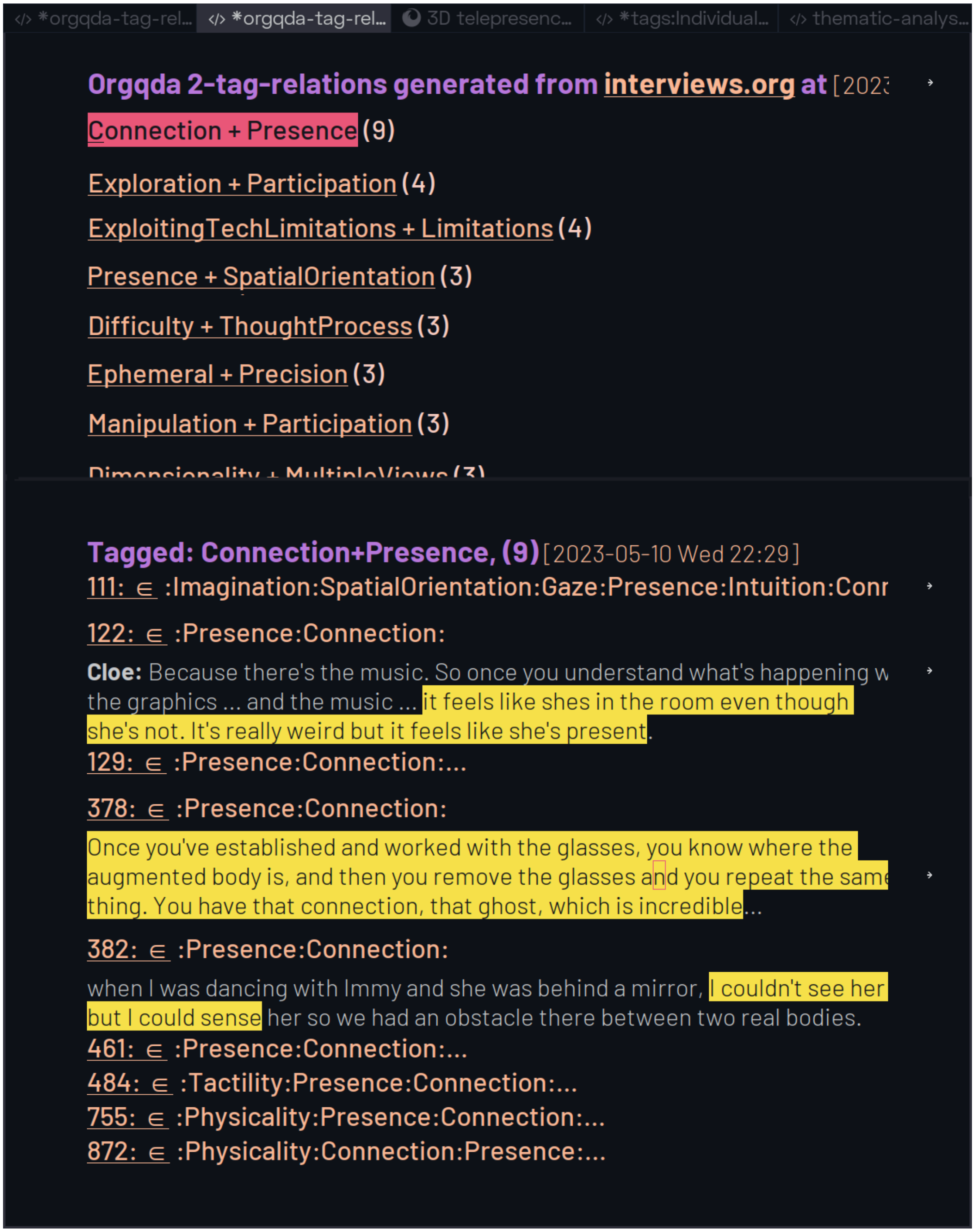
My approach involves a *collaborative* and *adaptive* development process, building and refining performance systems with a direct involvement of practitioners. This draws on *Participatory Design* — with system development occurring in tandem with actual performance creation to ensure the tools are fit-for-purpose and contribute to artistic intent.

Multi-modal data collection is used to construct a holistic record of the research journey. Performance documentation in the form of photos and videos, as well as in-development recordings, semi-structured interviews, group discussions and field notes present varied perspectives. This data is analysed using methods drawn from the *thematic analysis* process, where salient themes are identified and abstracted to construct deep insight.

Research findings are presented as both *artefact* and *theoretical* contributions. The artefact outputs — a fundamental part of both practice-based research and research through design — include a creative portfolio of mixed reality performances and their associated system designs, presented with context and analysis. The theoretical contributions are presented as a framework consisting of a set of system design heuristics and a collection of modes of interaction, derived from the iterative design and thematic analysis.

The methodological approach to my research strives to embody the intertwined nature of theory and practice in the context of mixed reality performance.

Figure 3.3. – *orgqda identifies code relations to help users interpret recurrent themes in the source material. The top window shows all two-code relations present in the selected datasets, and the bottom window shows all the excerpts coded with* both *'Connection' and 'Presence'.*



4. Augmented Performance with the AirSticks

4.1. A controller for multi-modal mixed reality

The works presented in this chapter are underpinned by a digital musical instrument, the *AirSticks* (Ilsar, 2018), and are the result of a fruitful collaboration with Alon Ilsar — the co-creator and principal performer of the *AirSticks*. Our partnership features prominently in this thesis over several works. Alon creates complex soundscapes that are visualised with real-time graphics systems developed by myself, making use of motion data generated through his gestural performance with the device.

The *AirSticks* have evolved over recent years, but in the iteration presented in this thesis, the *AirSticks* utilise the *Razer Hydra* virtual-reality gaming controller (shown in figure 4.1). Released in 2007, this device predates the wide array of modern motion controllers for virtual reality available today. As a result, they feature a distinctive technological design. The Hydra utilise electromagnetic radiation to track their position and orientation in space relative to a 'hub' placed in front of the user. The reliability of the tracking technology used in this device is notable in comparison to today's common inertial measurement (IMU) or camera-based designs for motion control. The *Razer Hydra* sample at 250 times a second,

Figure 4.1. – *The Razer Hydra virtual-reality gaming controllers.*

offer precision up to a millimetre, are not affected by environmental lighting conditions, do not lose tracking when occluded behind objects, and do not confuse their geometric origin in the world after prolonged use.

Despite the precision of the electromagnetic technology used in the Hydra, this solution was not adopted as a standard for handheld motion-tracked controllers moving forward. The Hyrda only have a practical tracking range of around one spherical metre from their base-station, which restricts their use in virtual-reality applications where greater mobility is often desired. Sixense — the company behind the technology — failed to mass-manufacture any product that worked at a greater distance (Robertson, 2017) . The Hydra's restricted operational zone is not an issue for the AirSticks however, and over the course of several years, Alon and his colleagues implemented and refined a software system that transforms these virtual-reality controllers into a configurable musical instrument that Alon was able to learn, practice and master.

The objective of my collaboration with Alon is to integrate mixed reality visualisation into his existing practice of electronic music performance with the *AirSticks*. A significant part of Alon's work focuses on creating 'mappings' of the instrument that integrate sound and movement into multi-modal gestures in a way that feels expressive to a performer, and appears unified to an audience.

Naturally, as I begun to work with Alon and his *AirSticks*, I too sought ways to tightly link his gesture into my own experimental visualisation systems. With the precision data generated by the *AirSticks*, we have been able to produce some tightly-integrated audio-visual work together that evokes strong correspondence between the physical movements of performance and the manipulation of audio-visual worlds.

The following sections in this chapter discuss three configurations of the AirSticks used for mixed reality performance. As presentations of visually-augmented music practice, this work features strong emphasis on the interplay of sound and visuals. Tight synchronicity between these elements — the concept of *multi-modal integration* — is presented as an important factor contributing to cohesive audio-visual composition and performance within mixed reality media, and serves as a common thread through the chronology of the productions.

Conventional video projectors are used to present the visual components by projecting onto a transparent theatre fabric known as *scrim*. Scrim is placed between the audience and the performers so that virtual environments appear superimposed on the physical performance environment.

Additionally, works in this chapter make use of depth cameras, and explore the potential they provide for fusing digital environments into physical performance space. After the introduction of these devices in this chapter, their use in my audio-visual productions continues throughout this thesis and is featured as a central piece of mixed reality technology throughout.

Through the integration of these technologies into music performance practice, this chapter presents case studies that offer insight into the potential mixed reality has to offer in presentation of novel live performance. My aim is that the presentation of these productions showcase the transformational nature of mixed reality as a medium, and hint towards a future where the boundary between physical and virtual is increasingly blurred. By embedding responsive virtualities, the immersive potential of live performance is elevated, and new dimensions of performer and audience engagement are revealed.

4.2. Trigger Happy Visualised

4.2.1. A gesture-controlled mixed reality show

The first work of this chapter is *Trigger Happy Visualised* — an audio-visual performance featuring the music of *Comatone and Foley* performed by Alon Ilsar, with an interactive visual environment programmed by myself. This production explores the interplay between sound, graphics and movement. Digital projections on a transparent scrim create a mixed reality stage environment that is controlled through Ilsar's live performance gestures.

This work explores several modes of interaction within a virtual audio-visual environment embedded into the physical stage. The performer plays on virtual instruments suspended between themselves and the audience using the *AirSticks*, directs abstract reconstructions of their body through motion, and performs on a drum-kit that is virtually augmented through integrated visualisations. This section elaborates on two of these interaction modes as they are presented in videos 4.1 and 4.2.

Trigger Happy Visualised has been performed at numerous festivals and events over the past five years, and has provided audiences with a view into a magical world of multi-modal mixed reality performance. A full recording of the hour-long show is presented in video 4.3.

► **Video 4.1** - *An excerpt from the scene* Computer Rain *shows augmentations based on depth camera input. (01:00)*

Watch at phd.matth.cc/trigger-happy/computer-rain.

► **Video 4.2** - *An excerpt from the scene* Heifen *demonstrates graphics that virtually augment a standard drum-kit (00:30).*

Watch at phd.matth.cc/trigger-happy/heifen.

► **Video 4.3** - *Full performance at the 2020* Melbourne Fringe Festival. *(57:10)*

Watch at phd.matth.cc/trigger-happy/melbourne-fringe.

Figure 4.2 – *The live performance is captured by a depth camera. Gestures transform the reconstructed avatar as it is projected in front of the performance area.*

Figure 4.3 – *The pose of the performer, and the physical gestures created are reflected in the virtual environment.*

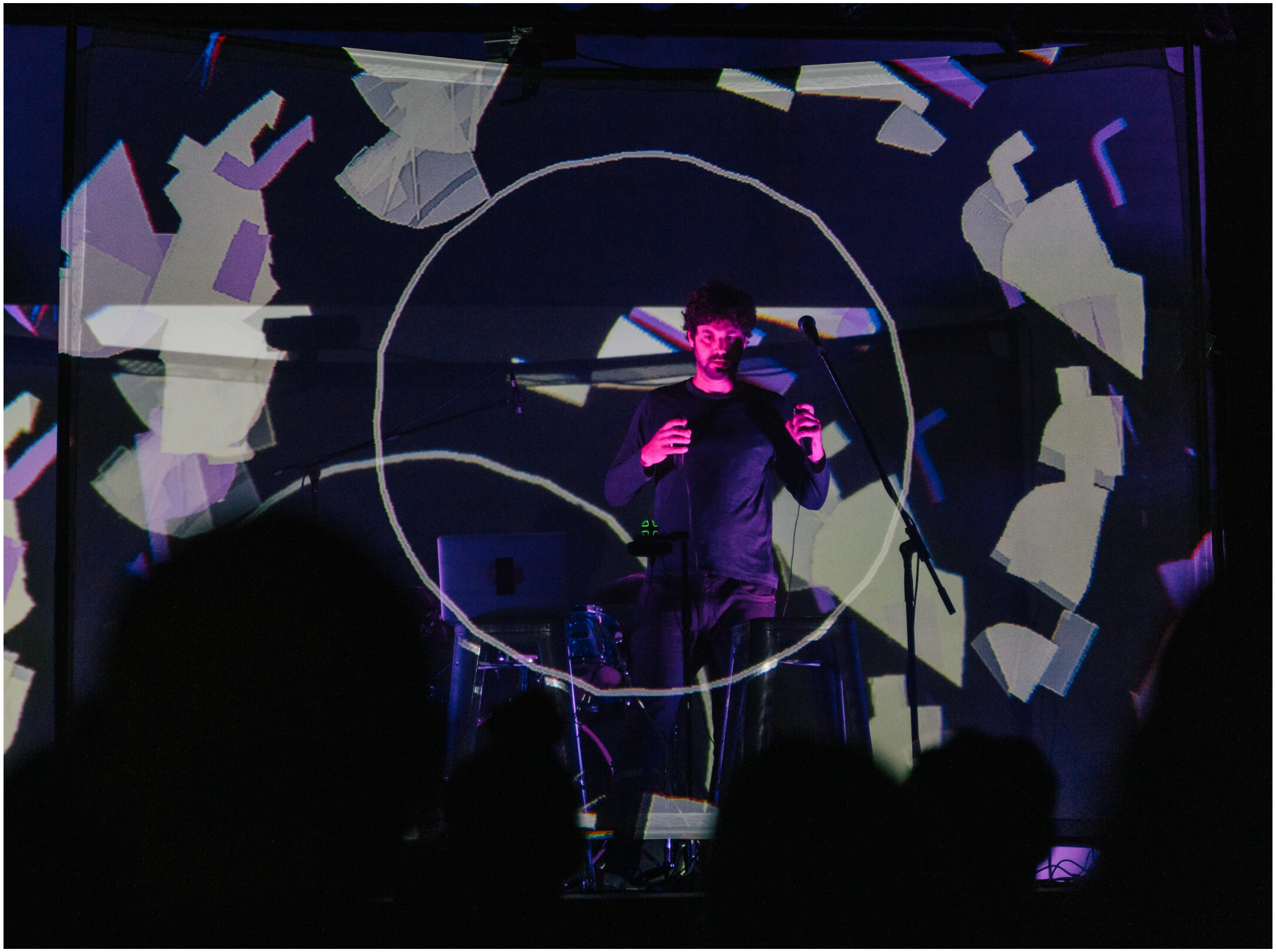


Figure 4.4 – *Interactive audio-visual objects are projected onto a pop-up scrim, and appear suspended in air between performer and audience.*

4.2.2. System design

Input data from the Razer Hydra is captured and analysed by a custom AirSticks mapping software developed by Mark Havryliv (Ilsar et al., 2013). The controller data is fed into a triggering subsystem that is able to infer 'strikes' through imaginary planes positioned in space around the performer. A strike generates a MIDI note-on message along with velocity, while raw data from the controller's position and orientation, along with button presses and analog sticks, are translated into MIDI continuous-control (CC) messages. Along with MIDI, the AirSticks mapping software simultaneously generates OSC messages that provide greater resolution than MIDI for subsystems that require it.

MIDI is sent into a range of synthesisers, sample-players and effects units inside the Ableton Live digital audio workstation[1] running on the same computer. This live interaction is combined with a prepared timeline that advances through backing-tracks for the music presented in the show, as well as pre-programmed MIDI sequences that map to pre-recorded aspects of the music.

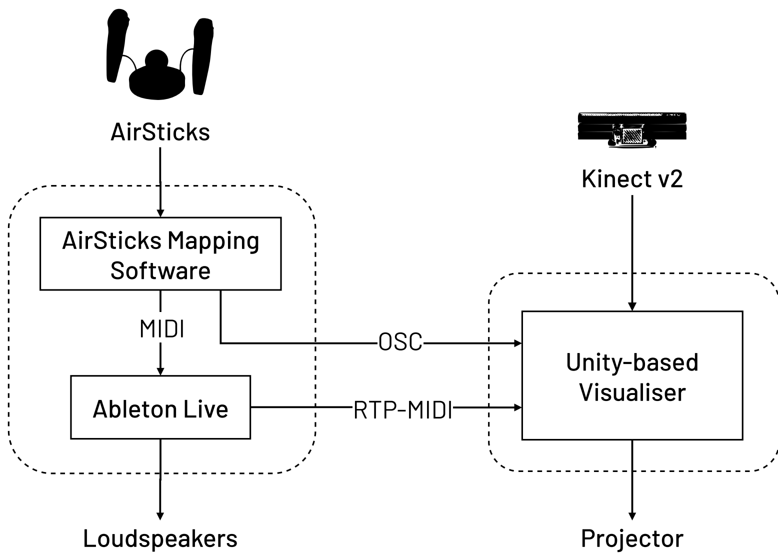
Simultaneously, a Microsoft Kinect v2 depth camera[2] is connected to a computer separate from the mapping and sound systems that handles the real-time rendering of graphics. The Kinect SDK is used inside a custom visualisation software developed using the Unity 3D engine[3], which renders several bespoke virtual environments throughout the performance. The depth camera is used to construct some visual components, while others are synthesised by independent generative graphics and particle systems.

1 https://web.archive.org/web/20201110193511/ableton.com/en/live

2 https://web.archive.org/web/20230130120318/learn.microsoft.com/en-us/windows/apps/design/devices/kinect-for-windows

3 https://web.archive.org/web/20190605231113/unity.com

Figure 4.5. – Trigger Happy Visualised *distributes tasks over two networked computers.*



MIDI from Ableton Live is sent over a local-area-network to the visualisation computer using the RTP-MIDI protocol[4]. This MIDI information is used to trigger visualisation components with note-on events, and is also used to change scenes and fire off pre-determined c ues. The OSC generated by the AirSticks is concurrently streamed into the visualisation software, where it is used to manipulate components of the virtual environment for gestures that require a higher precision than MIDI.

Audio from Ableton is sent to venue loudspeakers, and the graphics from the visualisation computer are transmitted to a projector pointed at the front of the stage. The projector casts images onto a transparent scrim positioned between the performer and the audience (visible in figure 4.4). The scrim — used not just in this work, but throughout many works presented in this thesis — facilitates superimposition of the virtual scene onto the stage to create a mixed reality

[4] https://datatracker.ietf.org/doc/rfc4695

performance space. It creates an MR illusion for the audience, but also provides a canvas in front of the performer that allows them to see and interact with virtual performance elements.

An outline of the system design is shown in figure 4.5.

4.2.3. Augmented physicality

By definition, mixed reality contains some blend of live physicality with live virtuality. If a virtual component of a scene is made responsive to a physical component or vice versa, the blend is more cohesive. *Trigger Happy Visualised* captures multiple sources of the physicality present on stage as inputs used to make the virtual environment responsive to the physical.

The most obvious physical input is in the motion control of the AirSticks, which manipulate audio-visual elements based on the performer's live pose. Many scenes in this production however, utilise additional interfaces between the synthesised virtuality and the stage. In these scenes, the physical presence of the performer and the performance area are amplified by interactive visualisations derived from the on-stage environment. The presentation of virtualities that are both derived from and used to transform the physical realm could be described as *augmented physicality* — where tangible, physical components are amplified through virtualities that elevate a material presence.

Two scenes in particular from *Trigger Happy Visualised* can be examined to further explore this concept. The use of a depth-camera in *Computer Rain* and a drum-kit in *Heifen* are both examples of mixed reality where virtual elements are informed by the physical environment in order to augment the physicality on stage.

Figure 4.6. – Computer Rain *uses a depth camera to influence the shape of a particle system.*

Computer Rain

The first example, *Computer Rain* disperses a swarm of illuminated particles among the performance area. The depth camera is pointed directly at the performer, and a 3D reconstruction of their body is fed into a simulation that drives the particles. The degree to which the particle system conforms to the shape of the performer's body is controlled by the vertical position of the right-hand controller (as shown in figure 4.6 and video 4.1). The horizontal position of the left-hand controller steers the trajectory of the particle system. As the performer shifts this controller left or right, the particles follow suit as if swept by an invisible wind.

The same movement allows the performer to play the main musical melody over a bed of cut-up rhythmic samples of a struggling computer's hard-drive — with pitch mapped from left to right. A strike with the right hand triggers a kick drum synth. The position and rotation of the controller adjusts timbre and panning, while delay and reverb are controlled by the heights of the left and right controllers, respectively.

Figure 4.7. – Heifen *binds virtual bursts of graphics to the physicality of the drum-kit.*

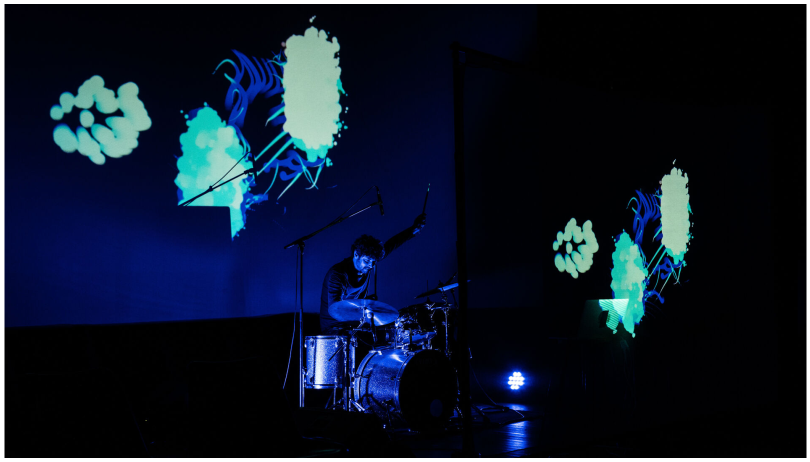
The manipulation of virtual elements so tightly integrated with the live physical form of the performer presents a vivid example of a system that enables augmented physicality. The performer's physical presence on stage is amplified by the use of the live depth stream as input for the visualisation. This mixed reality configuration exhibits a symbiotic relationship between the physical and the virtual.

Within the work's creative context, this relationship is used to present a manifestation of humanity's relationship with data. Oscillating between states of order and chaos, this audio-visual scene explores the tension between noise and clarity, both figuratively and literally.

Heifen

Heifen binds visualisations to the drum patterns played live on a standard drum-kit (as shown in figure 4.7 and video 4.2). Both the drum-kit and the performer's physical presence on stage are amplified through transient visual bursts as each drum is struck.

Figure 4.8. – *A seat on the side misses out on the desired spatialisation virtual elements in* Heifen.



An important factor in the success of this scene's augmentation of physicality is the effective localisation of the virtual elements onto the positions of each drum within the physical space. From the optimal viewing angle, reailty and virtuality is blurred with a high level of integration.

While the simulated elements appear strongly integrated, their interactivity is an illusion. It would be possible to achieve a comparable outcome that is genuinely interactive through the use of piezo microphones attached to the drums to trigger the visual bursts. However, since the music for this production was fully pre-composed and performed alongside a backing-track, a simpler solution was utilised whereby a prepared MIDI track of the drum pattern played through Ableton Live triggers the visualisation independently of the kit on stage.

While *Computer Rain* focuses primarily on using the performer's body through the depth camera as an input that binds physicality and virtuality, *Heifen* incorporates the tangible object of the drum-kit into the virtualisation. This allows the inanimate drums to appear as active participants in the mixed reality experience, extending the implementation of augmenting physicality to include the broader performance space, not just the performer.

When viewed off-axis, the mixed reality illusion in scenes like *Heifen* that feature specifically spatialised virtual elements can be compromised due to the nature of scrim-projection. As illustrated in figure 4.8, a viewer seated at a sharp angle off the projection surface doesn't experience the intended localisation of the drum-kit augmentations, reducing the degree of integration in comparison to a viewer seated parallel to the stage.

Though viewer perspective does still have a detrimental impact on the mixed reality experience, link between physicality and virtuality is more resilient in *Computer Rain*, as the depth camera is used to construct a partial re-implementation of the physical space. This maintains an embodiment of the performer's stage presence inside the virtuality even when viewed from off-axis.

4.3. Virtual Audio-Visual Instruments

4.3.1. Dynamic construction of a mixed reality

The previous section showcased bespoke audio-visual environments that combine live interaction with automation to render a mixed reaily performance for pre-written music. In contrast, this section presents the *AirSticks* as a controller for *standalone* audio-visual instruments that can be played indepent of any precomposition. These instruments build upon interactions from *Trigger Happy Visualised*, but remove the strict relationship with that production's musical timeline.

This work allows for versatile composition and improvisational performance through a set of three fully-interactive virtual instruments. Each instrument is manipulated using gestures from the AirSticks, and is visualised in the performance space by projection onto a transparent scrim. The instruments can be controlled simultaneously and sequenced. The result is a gestural performance system that enables the real-time construction of a multi-layered mixed reality scene.

The configuration of the AirSticks as discussed in this section was presented at the Georgia Institute of Technology's *Guthman New Musical Instrument Competition*, and at SIGGRAPH Asia's *Real Time Live!* in 2019.

Videos of these presentations are available at:

- ► **Video 4.4** - *The AirSticks at the* Guthman New Musical Instrument Competition *in 2019. (04:26)*

  Watch at phd.matth.cc/airsticks/guthman.

- ► **Video 4.5** - *The presentation of the AirSticks at SIGGRAPH Asia's* Real Time Live! *in 2019, which includes an explanation of the system design. (08:45)*

  Watch at phd.matth.cc/airsticks/siggraph.

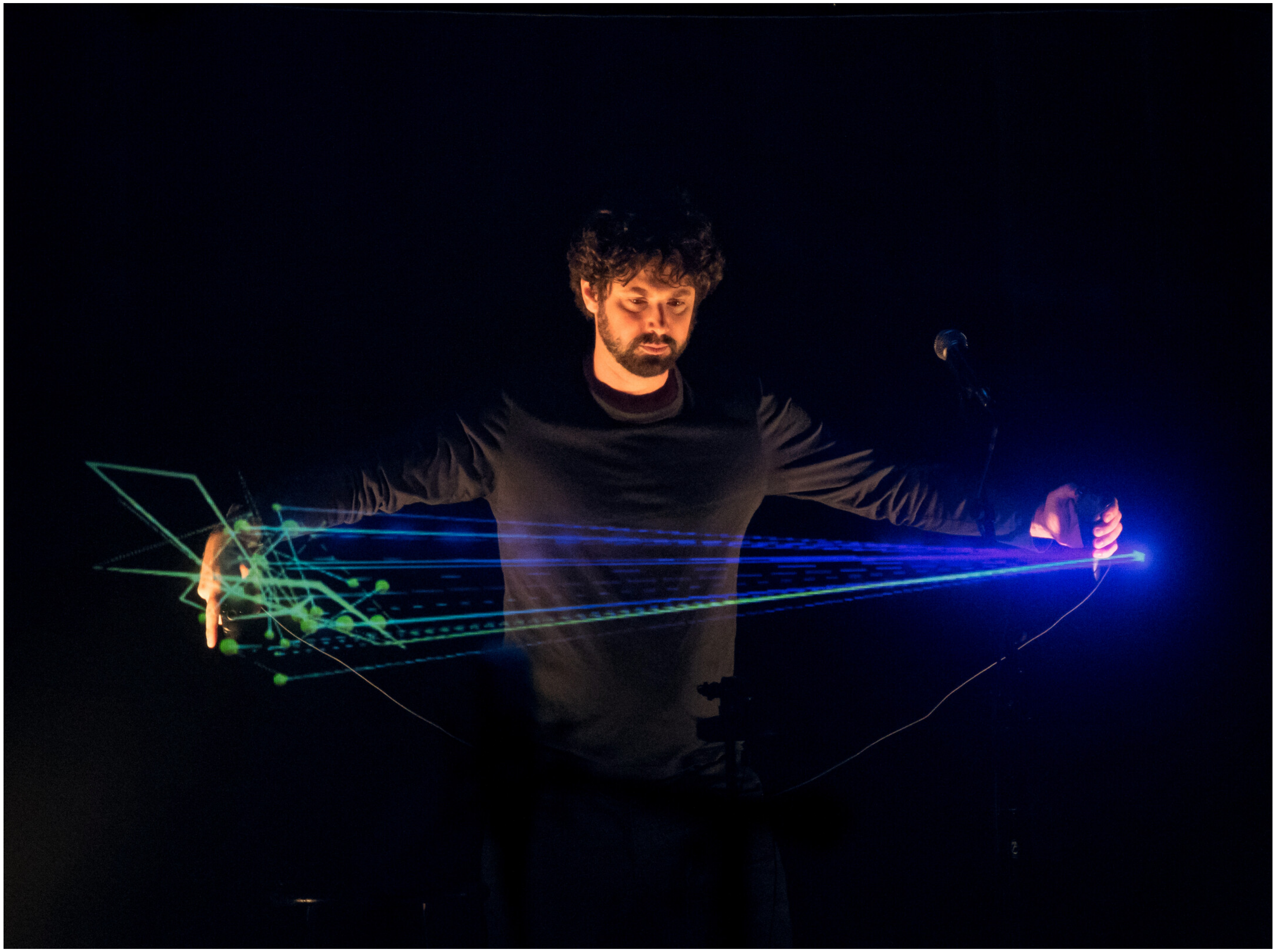


Figure 4.9 – The Rods *create an explosion of interconnected particles when the AirSticks are struck. Stretching them alters the behaviour of the particles and the timbre of the sound.*

Figure 4.10 – *Multiple audio-visual instruments are layered to create complex mixed reality compositions.*

Figure 4.11 – *Different configurations of each instrument can be programmed into presets to store variations.*

4.3.2. System design

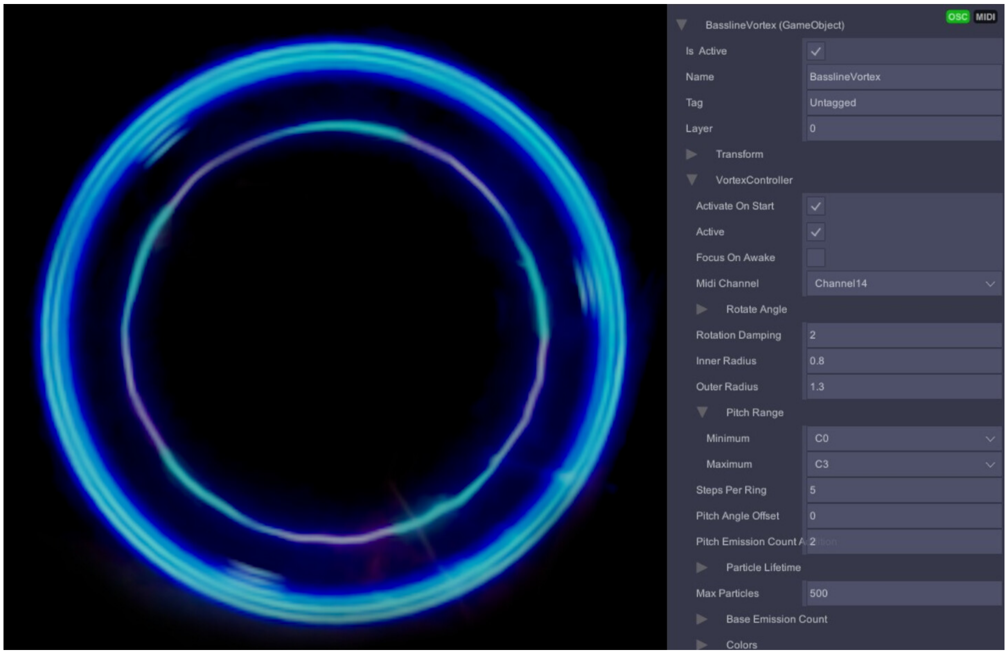
The technical infrastructure used for this work is similar to *Trigger Happy Visualised*. The AirSticks control Ableton Live on a Macbook Pro, as well as a Unity-based visualisation on a networked gaming computer — in this case a highly-portable Intel NUC. The audio is sent to loudspeakers and the video is projected onto the same scrim placed between the performer and the audience.

One key difference between this setup and the one used in *Trigger Happy Visualised* is the software user interface that is used to map the visual instruments to the incoming data from the AirSticks. With the previous system, all visual composition was done inside the Unity 3D Editor development environment. This time however, a custom user-interface that is overlaid on top of the visualisation at runtime was developed to provide focused control over the construction of audio-visual mappings.

This enhanced interface (shown in figure 4.12) is in essence a preset editor designed for configuring instances of the three virtual instruments. The controls of this interface allow graphics parameters to be easily mapped to incoming MIDI or OSC messages sent by the AirSticks. This user interface affords fine-tuned adjustments for visualisation behaviours and presentation, and allows raw data from the AirSticks to bind directly to these variables. At any point, the current state of these parameters can be saved to disk as a preset with save-and-restore functionality.

The performer and I engaged in co-design sessions where we collaboratively tweaked mappings in this editor against sound from Ableton Live in order to design convincingly integrated audio and visuals. When we achieved a result that aligned with each of our aesthetic goals, as well as the performer's needs for playability, the configuration was saved as a preset to be loaded at performance time.

Figure 4.12. – *The AirSticks mapping UI controls the behaviour of the visualisation.*



4.3.3. Harmony between sound, graphics and movement

This work focuses greatly on creating 'harmonic' relationships between the sonic, visual and gestural components of the performance. It engages in the practice of *audio-visual harmony* as described by computer animation pioneer John Whitney (1980) and extended by digital artists in the modern era (Alves, 2005). Audio-visual harmony concentrates on the integration of sound and image into a single entity, often generating both components from the same initial data source for more effective fusion (Fox, 2007; Ikeshiro, 2012). With the AirSticks, we introduce gesture into the concept of audio-visual harmony and attempt to bind all three constituents into one so that the AirSticks is perceived truly as an *audio-visual instrument* being controlled through performance gestures.

Figure 4.13. – The Rods *are a virtual mallet percussion instrument that feature distorted tones of a pentatonic scale arranged in space, visualised by an orbiting interconnected particle system.*

The human brain has the ability to combine multiple sensory inputs (e.g. sight and sound) into a single coherent event (Stein & Meredith, 1993). This process is referred to as *multi-modal integration*, and is contingent upon both temporal and spatial criteria. In essence, if sensory stimuli occur closely together in time and space, the brain is more likely to perceive them as a unified, singular event. In combining sound, graphics and movement with the AirSticks, we attempt to harness this phenomenon in the design of three virtual instruments suspended in front of the performer:

1. The Rods

The first instrument is named *The Rods*, pictured in figure 4.13. The Rods are designed to emulate interaction with a mallet percussion instrument. Five distinct regions are delineated in the performer's front space, each placed in different locations on the X and Z axes. Upon striking a virtual 'surface', each region produces a unique tone within a pentatonic scale. Accompanying the sonic response, each note's transient is visualised with a burst of particles, emphasising the impact

Figure 4.14. – *The innermost circle in these images is* The Ring*, a membranous circle that vibrates when struck to produce electronic drum sounds.*

of the performance gesture. The particles are interconnected between the left and right hands through elongated 'rods'. The specific notes struck determine the colours of these visual components.

The sound is sustained as long as the AirSticks maintain their position below the struck virtual surface, during which the particle visualisation continues to orbit around each controller. The absolute position of the controllers determine both the location of the visualisation, as well as the timbre of the sound, allowing the performer to stretch and pull each note within space to achieve a wide range of timbral variation. The behaviour of the visualisations is intrinsically linked to this sonic manipulation, and both the sound and graphics become more or less chaotic and noisy depending on the position of the performer's hands. As timbre intensifies, so do the rods and the particles that form the connections.

2. The Ring

The second instrument, titled *The Ring* is a membrane-like circle that surrounds the silhouette of the performer. It is the innermost circle pictured in figure 4.14. This circle can be 'hit' by one of the AirSticks through a strike to the air in front of the projection. A burst of angular noise emanates from where it is struck, while a similarly noisy percussion sample is sounded. As the audience views it, the performer can strike in different locations around the circumference of the ring to trigger a range of different audio-visual events.

Five virtual 'hit-boxes' for the AirSticks are placed around the circle, with each set to trigger the start of an electronic drum sound and a generative animation that draws from the shape and tone of the sound. After one of the ring's hit-boxes has been struck, a twist of the controller visually contracts The Ring's membrane, pulling it towards the strike point while engaging a band-pass filter, 'tightening' both the sound and graphics. The effect remains active until the user moves their hand away from the ring, causing a 'note-off'.

Buttons on the controllers are assigned to the task of recording a looping sequence on the ring. A button-press will put the ring into 'record mode', visualised by a change in colour to a neon-red glow to provide the performer and audience an indication of recording. With further button controls on the AirSticks, the user can start, stop and clear recordings, as well as overdub the sequence to build up complex rhythmic patterns. When record mode is deactivated, the ring returns to its original colour and continues playing the recorded drum sequence autonomously, freeing up the user's hands to play the other two instruments while the loop continues.

3. The Vortex

The third instrument, *The Vortex*, is an expressive arpeggiator visualised by spinning particles that radiate outwards from the performer. These particles leave trails to create rotating rings as the sequences run in an effort to visualise the cyclical workings of the arpeggiator. Unlike the previous instruments, *The Vortex* primarily operates through continuous control inputs as opposed to strikes in the air.

Figure 4.15. – *The* Vortex *ranges from simple to complex as instances are layered and manipulated.*

Each hand controls an arpeggiator for a different sound. The right hand is used to control a bass arpeggiator, and the left hand controls a higher, more percussive sound. To activate either arpeggiator, the performer holds down a trigger on the corresponding controller.

Moving the controller backwards and forwards in space alters how the arpeggiator climbs the chosen chord. As the notes in the arpeggiator climb, the diameter of the circle that the visualised particles travel on expands outward, resulting in dilating concentric rings.

Moving the controller left and right alters the timbre of the sound. For the bass, the horizontal movement alters the level of distortion, and for the percussive sound, the horizontal movement morphs its timbre from a hi-hat-like and completely non-tonal sound to one that sounds digitally melodic.

The analogue sticks on both controllers are used to change the rhythmic pattern played by each arpeggiator. Moving the analogue sticks left and right allows expressive control over rhythm with a mechanism that facilitates quick, immediate changes. This enables the performer to weave complex rhythmic textures in real time.

4.3.4. Practicalities of multi-modal integration

The aim of this work was to create harmonious and clear integration of sound, graphics and movement, although the projection-based approach to mixed reality used can make this multi-modal integration challenging. As previously noted, the effectiveness of multi-modal integration in practice is influenced by both temporal and spatial factors. We presented this mixed reality configuration across dozens of venues over years, and found that performance venues can fight against both temporal and spatial requirements in our attempts at audio-visual synchronicity with the instruments.

The construct of the *temporal binding window* helps highlight a potential downside we discovered when relying on a projector to present the integrated audio-visual objects of this project. Separate audio and visual stimuli are perceived as one when these stimuli are presented temporally close to each other (Alais et al., 2010). Neurological studies have attempted to quantify the temporal range that multi-modal stimuli must lie within to be considered integrated (Dixon & Spitz, 1980; Stevenson et al., 2012). The temporal binding window refers to this range, within which events across sensory modalities are likely to be perceived as a single event. A strike of the AirSticks generates a sonic event and a graphic event in reaction to the movement. For a participant (either user or spectator) to believe the experience as an integrated audio-visual instrument, the events across both senses need to occur within the participants temporal binding window. Stevenson et al. (2012) found that the mean size of the window present in the participants of their study was 215 milliseconds. Given that this is the *mean* time window we have to integrate sound and vision into one, then an audience could contain many people for which this window is smaller.

For the AirSticks to create an audio-visual gesture, the system pipeline must:

1. Process the performer's movement to generate an event
2. Communicate this event to connected audio and visual subsystems
3. Generate audio and send it to speakers
4. Compute the visualisation and send it to a projector

The virtual reality controllers that the AirSticks are based on — the Razer Hyrda — have a latency between 4 and 10 milliseconds (Basu et al., 2012). Communication of signals from the AirSticks over the local network of connected computers takes no more than 1 millisecond. To generate audio, the chain from MIDI input to a digital instrument to the audio interface could take up to 30 milliseconds. Given this, we can make a reasonable conclusion that both the movement gesture and a *sound* produced by the AirSticks will fall within most people's temporal binding window. In addition, because we use the same controllers, computer hardware and audio interface for each performance, the latency in converting movement to sound is constant through all presentations of the performance system.

For the visualisation however, achieving a low latency can be more challenging. It is not unreasonable to expect a frame of visualisation to take 16 milliseconds to render, which would give us ample capacity for audio-visual temporal binding — if it were not for the projector. Over our performances in many different venues, it became clear that projector selection can have a significant impact on the latency from gesture to visualisation.

The signal-to-visual delay of a projector is the critical factor. This parameter is communicated as *input lag* in projector specifications. Large-format and installation projectors used in big venues often focus on scale and brightness rather than latency, and can exhibit considerable input lag in the realm of hundreds of milliseconds. This is not a top priority for manufacturers of large-format projectors, because a few hundred milliseconds is generally suitable for most other applications.

In addition to the input lag present in the projector itself, the video signal chain can introduce even more delay. Video signals from the PC to the projector may require very long cable runs that introduce delay; they may be amplified or sent through HDMI matrices, capture cards and converters; they could be converted to Ethernet before being converted back. All of this together can seriously impact the time it takes for a physical gesture to be expressed in the virtual world.

Video 4.4 shows a prime example where the model of projector had a high input-lag, resulting in a clear impact on the audio-visual integration. From the very first strike it is noticeable that the virtual elements floating in front of the performer lag both the gesture and sound by a considerable amount.

For smaller scale presentations, where relatively less projection size or brightness is required, *gaming projectors* may be suitable to ensure a low amount of input lag. These have a dedicated mode that reduces input lag as much as possible to aid with a player's reaction time in multiplayer gaming, though projectors with this feature often require *geometric correction* to be disabled in this mode. Geometric correction is a software feature that allows a projector to adjust the image to compensate for the shape and position of the projection surface. It is hard to imagine a presentation in a live performance setting that could forego this feature, which makes the subset of gaming projectors that disable geometric correction in low input lag mode also unsuitable.

If possible, practitioners should use their own projector or make sure one that is being provided features low latency, however it is not always possible to choose the projector when playing different venues unless the performance has a commercial budget. In addition, cable runs and conversions to and from video signal formats should be minimised as much as possible, though again this can't always be achieved. Unfortunately, very tightly bound temporal synchronicity with the goal of creating integrated audio-visual instruments is diminished with many projection configurations. Practitioners utilising projection-based mixed reality may have more success designing interactions that do not require high temporal precision. Slow gestures that are more forgiving to the input lag may present better across a variety of projectors in comparison to gestures that accentuate transient-focused 'strikes' like the Rods and Ring do.

Depending on the design of the virtual instrument, spatial synchrony may be as important as timing. This was discussed briefly in section 4.2.3 on *Heifen* in *Trigger Happy Visualised*, though the rods here are another example of a visualisation that maps physicality and virtuality in a one-to-one spatial relationship. The rods appear just in-front of the AirSticks, tracking and following the physical location of the performer's hands. Notice the rods in figures 4.14 and 4.13 that appear to be shooting straight out from the handheld controllers. With this kind of visualisation, the accuracy of the instrument's graphical representation in space is a consideration. However the ability to manage this factor is often out of our control, again due to the projection-based presentation method chosen to visualise the mixed reality.

A characteristic of projection-based mixed reality is the flat plane of the projection surface. In contrast to the three-dimensional experience offered by headset-based mixed reality (explored in section 5.4), the scrim used for the AirSticks ensures that localised visual events, like the Rods, work to their full effect only when the audience is located front-on to the stage. The Rods only appear as spatially mapped to the performer's hands from within a limited range of angles approaching perpendicularity to the projection surface. Through the spatialisation of the Rods, we attempt to tightly integrate the performer's physical pose with the virtual graphics. People are more likely to report stimuli as integrated when their spatial position is coupled (Zampini et al., 2005), though this is hard to achieve uniformly when audience position dictates the ability to accurately spatialise the virtual elements. From the front, the virtuality would appear aligned to the body, but that experience is lost at steep angles. For many performance venues, it is inevitable that audience sight lines will be spread across a wide range of non-ideal angles, which fights against attempts at tight spatial integration of physical and virtual worlds when using projection for mixed reality.

4.4. Computer Storm

4.4.1. Bringing in the crowd

Computer Storm examines our existence as streams of data, and intimates a complex relationship with surveilence by directing depth cameras at the audience in addition to the stage. It was presented at the 2020 conference on *Tangible, Embedded and Embodied Interaction* (Ilsar & Hughes, 2020).

This audio-visual piece is a spiritual successor to *Computer Rain* (4.2.3). Like its predecessor, it affords the performer control over a field of particles with the AirSticks, and through the use of depth cameras, he is able to oscillate between noise and human form.

An interactive virtual environment is projected between the performer and the crowd that contains reconstructions of them both. As Ilsar navigates the digital landscape, audience-members become a part of the performance. Their bodies and interactions influence both the visuals and the spatial soundscape. This piece aims to dissolve the distinction between spectator and participant, raising questions on personal space and the implications of living in an increasingly connected world.

► **Video 4.6** - *Development preview performance for* Computer Storm *at UTS. (06:53)*
Watch at phd.matth.cc/computer-storm/development.

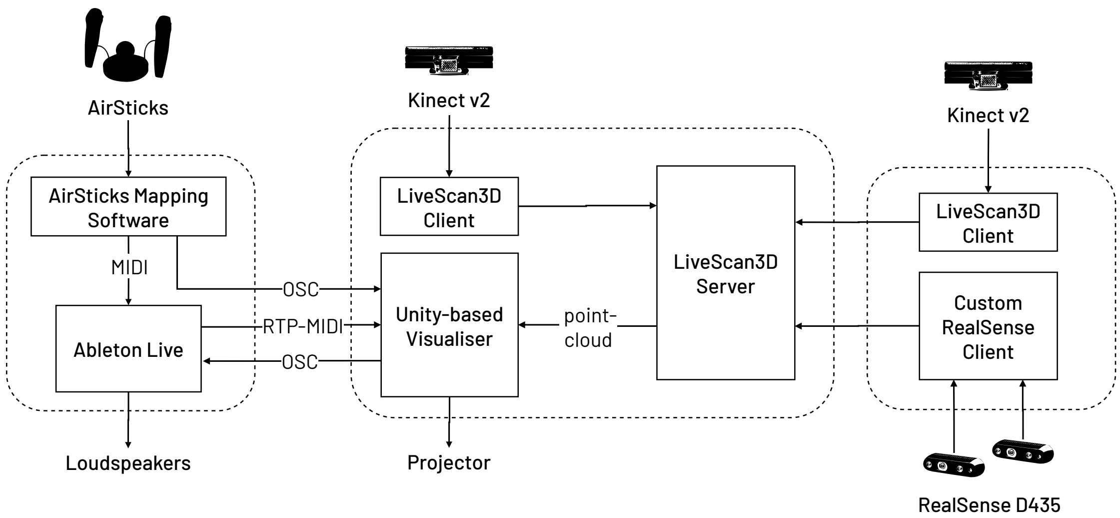
► **Video 4.7** - *A performance of* Computer Storm *with a view of the crowd at TEI 2020. (07:03)*
Watch at phd.matth.cc/computer-storm/tei.

Figure 4.16 – *The performer uses the AirSticks to nagivate a camera through a virtual reconstruction of the audience.*

Figure 4.17 – *The performer uses the AirSticks to nagivate a camera through a virtual reconstruction of the audience.*

Figure 4.18 – *The performeruses the AirSticks to nagivate a camera through a virtual reconstruction of the audience.*

Figure 4.19. – Computer Storm's *distributed system design.*

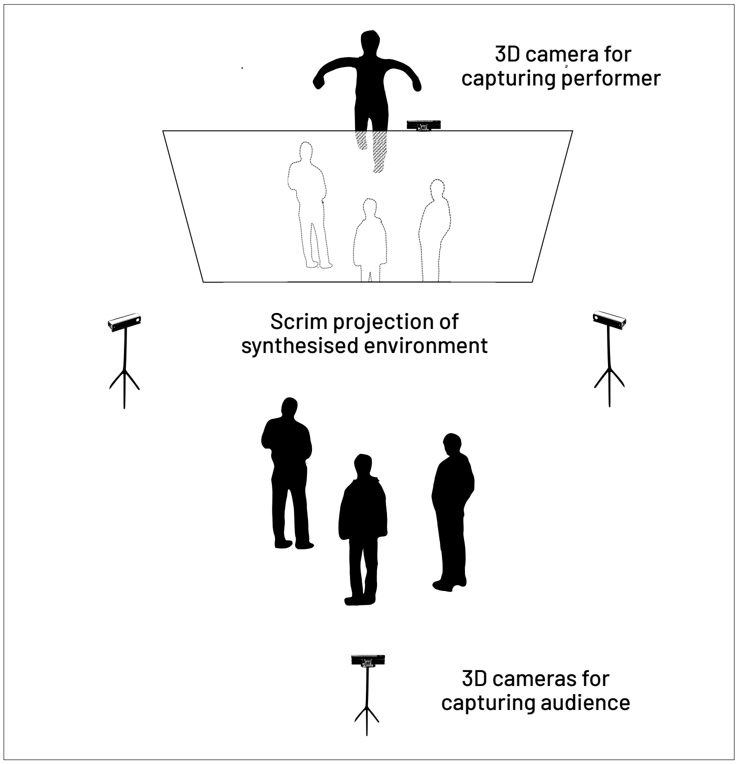


4.4.2. System design

Of the projects presented so far in this chapter, *Computer Storm* is the first to expand the area captured by depth cameras beyond the boundaries of the stage. A multi-camera setup is directed at the environment of both the performer and the audience to reproduce the entire performance space in a virtual scene that can be navigated and manipulated in real-time. To complement the comprehensive coverage of the cameras, the audio is presented in quadraphonic surround sound. The audio-visual scene is controlled through movement using the AirSticks, and like the other AirSticks projects discussed so far, the virtual world is projected onto a scrim positioned between the stage and the crowd.

Three network-connected computers are used in total. One PC hosts the AirSticks mapping software as well as Ableton Live for generating audio. A second PC hosts the three audience-facing cameras. The third PC hosts the performer's camera, a LiveScan3D server for synthesising the point-cloud, as well as a Unity-based visualisation server that generates the graphics. This distributed network of components is illustrated in figure 4.19.

Figure 4.20. – Computer Storm's *teleimmersive configuration.*



Depth cameras

LiveScan3D (Kowalski et al., 2015) is at the core of this setup. This software synthesises point-clouds captured from multiple depth sensors into a single, volumetric scene. Four depth cameras work in tandem to capture the entire performance environment, with one Microsoft Kinect v2 placed in front of the performer, two Intel RealSense D435s positioned in front of the audience on either side, and one more Kinect v2 placed behind the audience. This configuration is illustrated in figure 4.20.

The D435 cameras have the advantage of being able to connect to the same host computer simultaneously, in comparison to the Kinect v2 which require one computer per camera. A comparison between these cameras is explored in more detail in section 6.2.2. This simultaneous connection capability allows only two computers to be required to host cameras for this performance, though

In order to integrate the RealSense D435s into the LiveScan3D system which only supports Kinect cameras, a custom version of the open-source Intel RealSense Viewer tool[5] was developed with added network streaming capabilities. In this custom application, depth and colour frames from the RealSense cameras are packed and sent using the protocol required by the LiveScan3D server.

Despite the increased resolution of the D435 cameras compared to the Kinect v2s, the D435 sensors struggle when capturing subjects over 1.5 metres away. The initial plan was to use only three cameras in *Computer Storm* — one for the performer and two positioned on the sides of the audience, though the second Kinect camera at the rear needed to be introduced to combat the D435s short practical range. The range deficiency however, results in relatively sparse point-clouds coming from the D435s compared to the density that the Kinect v2 is capable of when positioned at a reasonable distance for capturing the audience.

As the Kinect stream is more dense than the D435s, the synthesised point-cloud generated by LiveScan3D using these two camera models generates a visualisation that emphasises the Kinect cameras more than the D435s. For *Computer Storm*, this potential flaw ended up harmonising well with the themes of the piece. As the performer navigated the virtual camera around the 3D form of the audience members, the backs of people (as captured by the Kinect) appeared far more clear

[5] https://web.archive.org/web/20220516102019/github.com/IntelRealSense/librealsense/tree/master/tools/realsense-viewer

than their front. This added a kind of 'anonymisation' to the audience members, resonating with with the piece's exploration of surveillance and our relationship with technology.

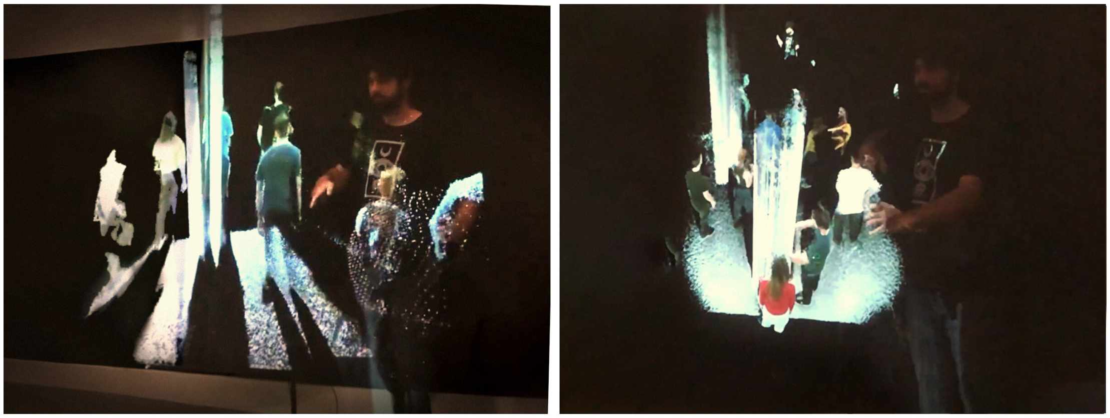
The Kinect v2 directed at the crowd however did present its own challenge, mostly due to its limited capabilities for controlling the behaviour of its sensor. We found ourselves wrestling with the auto-exposure feature of the camera. In dark settings, the Kinect v2 automatically switches to a 15 frames-per-second mode in order to let more light into its colour sensor. This doesn't pose a problem in bright rehearsal environments or for well-lit performers on stage. However, it can significantly affect the capture rate of audience areas, which are typically less lit than the performer. For our presentation at TEI, the audience was captured at half the frame-rate of the camera pointing towards the performer. This could have been detrimental to the piece, however the effect was barely noticed due to the relatively static nature of the audience during this performance (a detail discussed more in section 4.4.4).

The Drone

Computer Storm's main mode of interaction for the performer is in the navigation of the audio-visual environment. This space is navigated by OSC signals sent over a network from the AirSticks.

The left-hand controller orients the view into the visualisation by manipulating a virtual camera, shown in figure 4.21. The right-hand controls various noisy simulation behaviours applied to the particle system. The virtual camera's horizontal and vertical location can be altered by moving the left AirStick in any direction. The camera flies about the scene as if the performer is operating a *drone* in the virtual space. This camera drone can be flown anywhere in the reconstructed live

Figure 4.21. – *The performer orients the virtual scene in* Computer Storm *by positioning his left-hand controller. Simultaneously, he positions a 'vacuum' in the virtualised crowd.*



environment, and by pressing the left or right trigger on the AirSticks, the drone can orbit around its focal point. The following video shows a prototype of the drone-navigation interaction mode and its integrated sound design.

► **Video 4.8** - *Gestural interaction in* Computer Storm *enables control of a virtual camera drone. Be cautious with sound levels if wearing headphones. (01:12)*
Watch at phd.matth.cc/computer-storm/control.

This 3D camera drone makes it possible to view the real-time digital reconstruction of the physical scene through unlimited viewpoints. This includes orientations that would be impossible for a physical camera to occupy: hovering above the audience, sweeping among peoples feet, or gazing from inside someones head.

The graphics and the audio in this piece are integrated with many of the AirSticks' inputs configured to affect both modalities simultaneously. Moving the position and orientation of the camera drone in the virtual world alters the spatialisation and aesthetics of the surround-sound coming from Ableton Live. Moving the drone to a particular area of the scene focuses the sounds out of the loudspeakers into that space. Orbiting the camera results in sounds swirling around the crowd, and zooming in close into the personal space of the audience members produces intimate and closely-miked breathing sounds to evoke a sense of uninvited discomfort.

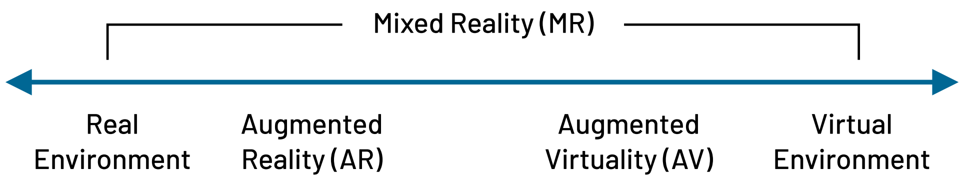
The Unity simulation also measures the point-cloud it receives from LiveScan3D, and the resulting values can send to Ableton over OSC to affect sound generation. Pressing a button on one of the AirSticks engages a 'vacuum' that animates the point-cloud upward as a beam of light. The performer uses the same controls as the camera to control this vacuum. If the simulation detects that an audience member is present in the space being sucked up by the vacuum, a rhythmic sequence is activated. If the audience leave the zone being activated, or the performer shifts their focus, then the sequence stops. *Computer Storm* builds to its climax as a few vacuums are locked in place among the crowd to generate a set of interactive counter-rhythms.

The objective when developing *Computer Storm* was to use depth cameras around the audience to explore the physicality of a wider mixed reality performance space. The end result is a piece that encourages engagement and attempts to draw the audience into participation through their live presence in the teleimmersive scene.

4.4.3. An Augmented Virtuality

Chapter 2.1 of this thesis mentioned the *virtuality continuum* as described by Milgram and Kishino (1994). This spectrum of mixed reality is pictured again below.

Figure 4.22. – *Milgram and Kishino's virtuality continuum.*



The previous AirSticks scenes discussed in the three instruments and *Trigger Happy Visualised* sections all create augmented reality environments. They superimpose virtual elements onto a physical environment with the goal of blending the two worlds into one. The spatial registration of virtual elements in the physical scene is a defining part of augmented reality, so some might say these works projected in 2D onto a flat surface do not satisfy its definition. To that I would respond that the virtual elements are still registered into a 3D simulation, and are manipulated in response to the physical 3D spatialisation of the AirSticks, the performer, and the drum-kit. Even if these works are not canonical AR, they at least sit toward the left side of the virtuality continuum above.

In contrast, *Computer Storm* does not attempt to present virtual objects as if they are part of the physical environment of the stage. Instead, it presents a virtual environment containing a reconstruction of the physical environment, and it does so as a visualisation that is not attempting to embed itself into the physicality of the stage.

This places *Computer Storm* further to the right on the virtuality continuum as an example of *augmented virtuality*. This piece in particular explores the augmented virtuality concept of *teleimmersion* — the embedding of *people* into a distinct virtual environment (Kurillo & Bajcsy, 2013). As a teleimmersive performance, *Computer Rain* is able to investigate the notions of participation and presence inside a virtual performance space.

4.4.4. Participation

Primarily, the development of *Computer Storm* served as a platform to consider audience participation within the scope of audio-visual performances. Although we barely touched the surface exploring the modes of interaction available to a performance like this, the creation process, presentation, and subsequent discussions with participants provided valuable insights into the dynamics of audience involvement within the medium of teleimmersive performance.

Some of the most entertaining moments for the audience-members were when they became consciously aware that they were not just spectators, but active participants.

> … the moment I realised that it was really 'live' … when it zoomed out, and we were all these colourful blobs … I could read the back of my shirt [and] I was like "oh damn!" … I felt like those moments stood out. — R

Initially the performer manipulates an image of themself on the scrim, but soon after pulls the virtual camera away and into the crowd. The transition from passive observer to active participant is not something people expect when they attend a performance.

> If we're watching, [we accept] 'being an audience', instead of being part of it. At first I didn't consider we were going to be on that screen at all. So we've taken on that role of audience, but when it was taken back it was a bit like "Am I watching? Or am I participating?" — I

This surprise shift in role is exploitable particularly in a performance context. The preconceived notion of what activity an attendee might engage in when they arrive to a music performance is markedly different from environments like galleries or museums, where an audience might be more likely to expect an interactive experience.

> When an audience is there to see a performance ... [they're] like "I'm in an audience and I'm just watching" ... but if it was ... in an exhibition or in a space where people were wandering around interacting with things ... you'd get a completely different reaction. — E

> If you make it an installation then you're not questioning the audience. you're telling them to interact. Whereas here, you come in going ... "I'm just watching" ... and that to me is the point of this. That's the challenge. — Performer

The vacuums provided an virtual audio-visual element that the crowd was able to influence themselves through their position in the room, but this mechanic elicited very little physical engagement for a number of reasons.

Complexity of the visualisation was one factor that inhibited participation. If the graphics overwhelm the participants, it might stun them into acting more observant than participatory.

> I've never seen anything like that ... I don't understand the tech of it ... so I was just starting at the screen. — I

The almost constant movement of the camera drone was another detail that caused restraint from the audience. To incite movement from the audience, it could be more effective to maintain static or only slowly-moving camera states.

> I think with so much movement you don't know when you should move yourself. — J

On reflection, it seems naive to think that we could promote a playful audience movement while the work's voyeuristic theme makes patrons uncomfortable. This theme characterised by the drone successfully induced an anxiety in the participants — at odds with kinetic participation.

> I wanted to move but I felt very self conscious ... because we're under surveillance. — I

The transformation of audience members from observers to contributors through mixed reality must be well-managed. Real-time capture of an audience into a virtual environment can offer a highly engaging experience, but abrupt changes or overwhelming stimuli can hinder participation. Future works should experiment with the balance of immersion and audience comfort.

4.4.5. Presence

The physical experience of *Computer Storm* extends beyond the active engagement of participants with the virtual environment. A passively experienced sense of physicality can be induced by the virtualities of the scene. A perceived *presence* of the drone inside the room was described by many participants. The virtual construct was able to elicit a quasi-physical response.

Even though audience members are watching themselves from an external perspective, they still feel the presence of the drone as if it was navigating the physical room. The presence of the drone is transported outside the display and into personal space, even though its not at all visualised.

> I … had that feeling like … it's definitely moving amongst us. — E

The drone's presence was enhanced by tight integration between the sound and the visuals. The spatial link between the surround-sound placement of the audio with the position of the camera drone within the projected scene further embedded the virtual drone in the audience's physicality.

> It felt like there was … an entity moving around, especially with the sound of the breath. — J

A real intimacy to this presence is demonstrated through the audio-visual integration of breath samples and camera zoom. As the virtual camera zooms closer into the crowd, the sound of breathing is focused and intensified.

> I actually felt like something was moving around us … when the sound was minimal. — D

The drone pilot interaction mode in the virtual space allows a presence in the physical space to be embodied by the performer. The control that the manipulator of a teleimmersive environment has over the psycho-physical perceptions of participants is an incredibly evocative power. And it becomes more powerful through multi-modal integration. This intense connection between the performer and the audience could have profound implications for creatively shaping future teleimmersive or participatory experiences like *Computer Storm.*

5. Telepresent Performance

5.1. Making connection

Computer Storm was one of my first meaningful explorations into the inclusion of audience participation in a mixed reality work. It had been my plan to continue down this path and further explore the integration of mixed reality into audience environments at live music performances, and the potential for this to engage people in novel ways. I had developed some comfort and experience with systems for producing mixed reality visualisations of live music with the AirSticks projects, and I had been creating my own system for audio-visual composition specifically dedicated to a live music context.

The first public presentation of *Computer Storm* was performed at the TEI conference in February of 2020. Less than a month later, my home state began to restrict attendance of public events due to the arrival of the pandemic. By the end of March all residents were mandated by the government to stay inside their homes, and it became very clear that a focus on conventional crowd interaction for any upcoming work was out of the question. The live music scene shut down for a year and a half, and when it opened back up, we faced another six months of restricted access and bans on standing crowds and dancing in venues. To continue my practice-based study, it became necessary to adapt to the new circumstances.

The concepts and technologies explored in the previous works were not made redundant but required recontextualisation. In this section titled *Telepresent Performance*, three new works are presented that repurpose and extend the technologies previously used for augmented live performance into productions that incorporate the concept of *remote participation*. The desire to overcome geographic boundaries and sanctioned separation led to a focus on *teleimmersive* and *telepresent* configurations of mixed reality performance.

These works explore the possibilities of the virtuality continuum in a new light. They evoke a sense of presence for performers within spaces that are virtual or physically remote. To connect over distance, the following productions make use of depth cameras and reprojection, and delve deeper into both virtual space and augmented reality.

5.2. AirStream

5.2.1. Creating a composite space

AirStream is an AirSticks duet that bridges the physical distance between two performers through an internet-connected teleimmersive environment. Each performer and their immediate physical environment is captured and streamed into a virtual space to form a composite living room — from one half in Sydney and the other in Melbourne. Both performers are reconstructed as live point-clouds in the virtual room, where they simultaneously manipulate sound and graphics through gesture and pose.

The extended pandemic lock-downs and inter-state travel restrictions in our home cities forced us to explore new methods of remote performance and collaboration.

Performance within teleimmersive environments like this highlights the potential of technology to foster human connection even in times of forced separation.

AirStream was presented at the conference on *New Interfaces for Musical Expression* (NIME) in June 2021, as well as at *Audio Mostly* in September 2021. The full performance recording is available below.

► **Video 5.1** - *The* AirStream *performance as presented at NIME 2021. (08:34)*
Watch at phd.matth.cc/airstream.

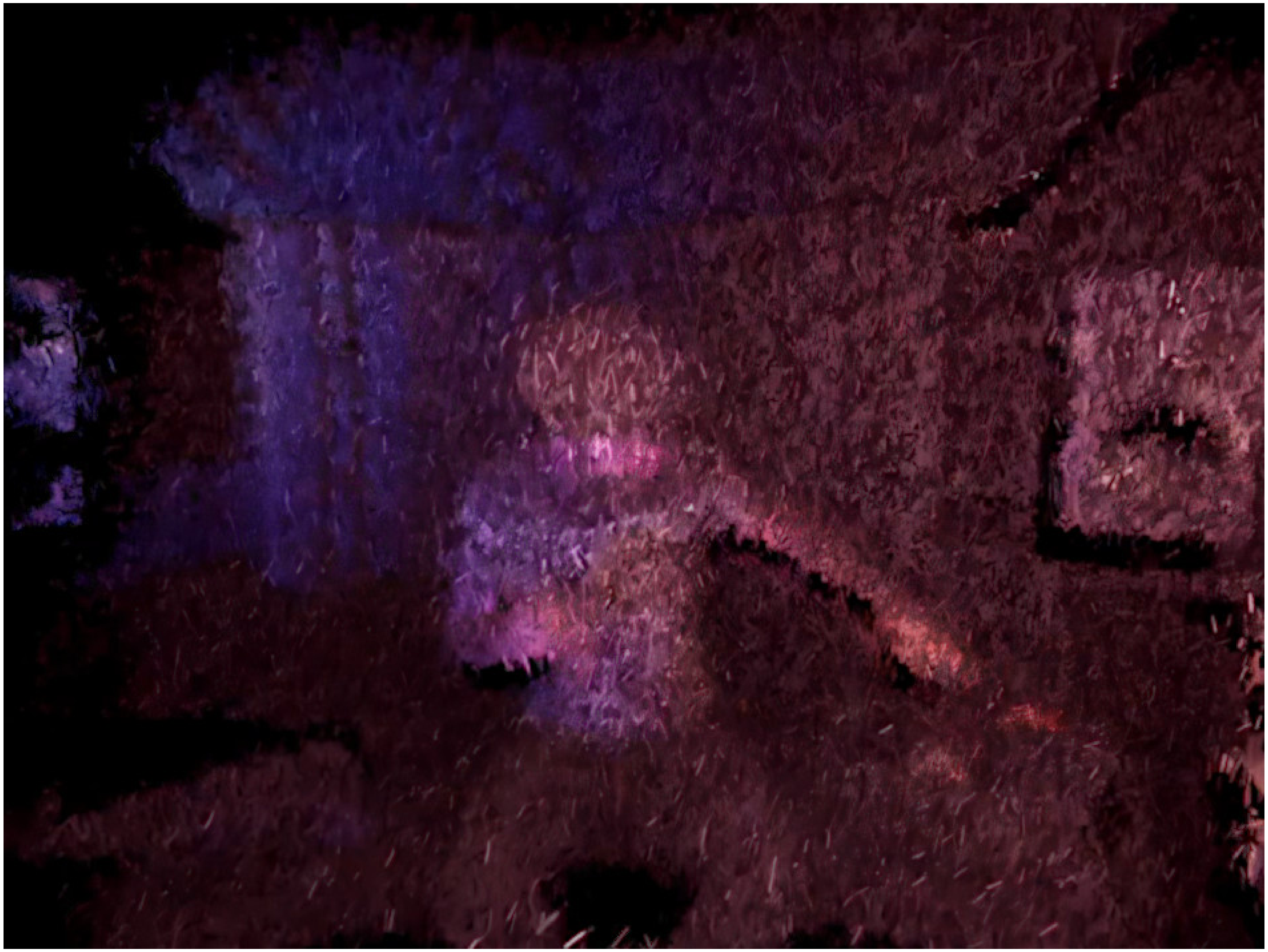


Figure 5.1 – AirStream *merges the rooms of two remote performers into a single audio-visual environment.*

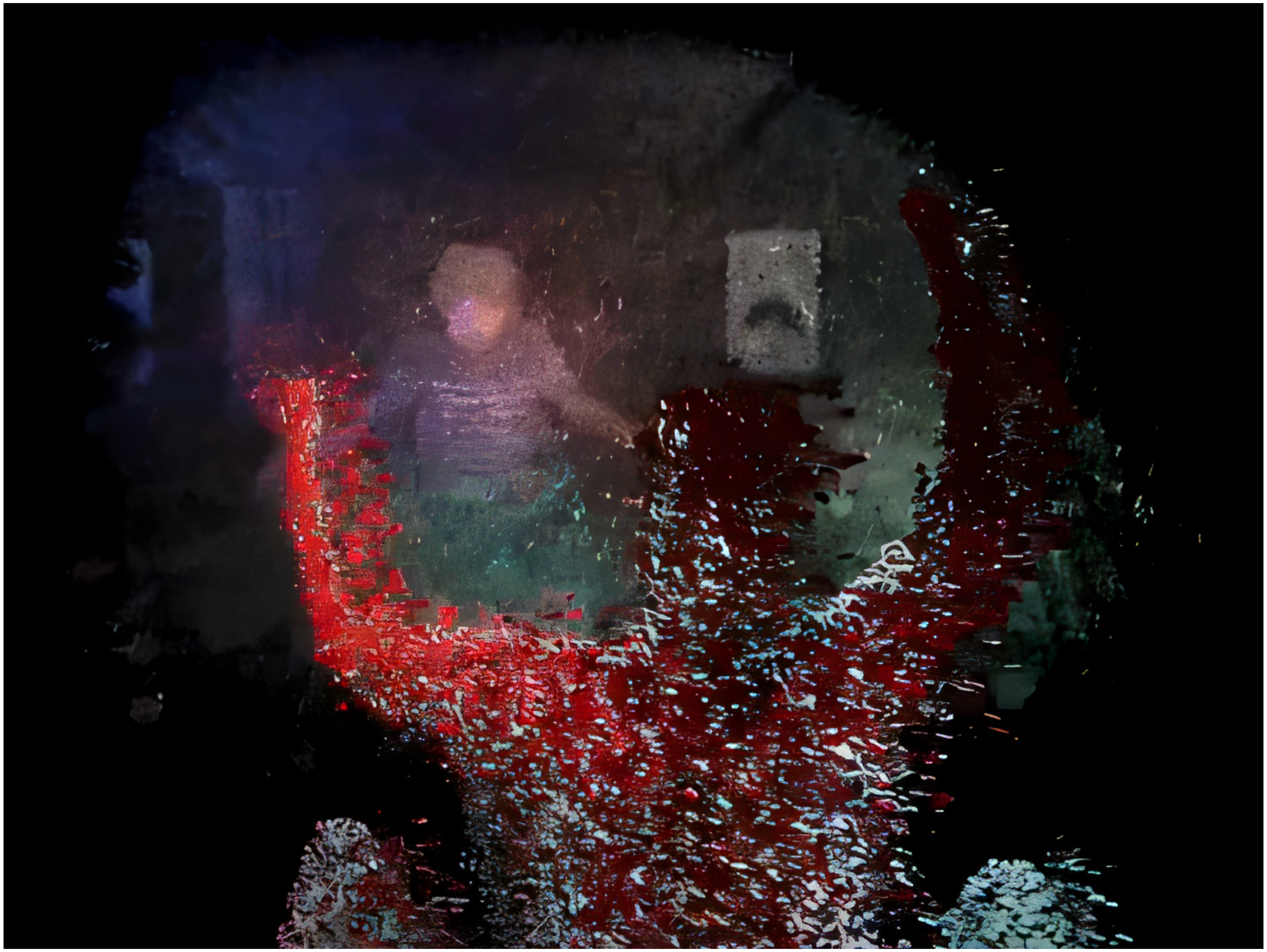


Figure 5.2 – AirStream *merges the rooms of two remote performers into a single audio-visual environment.*

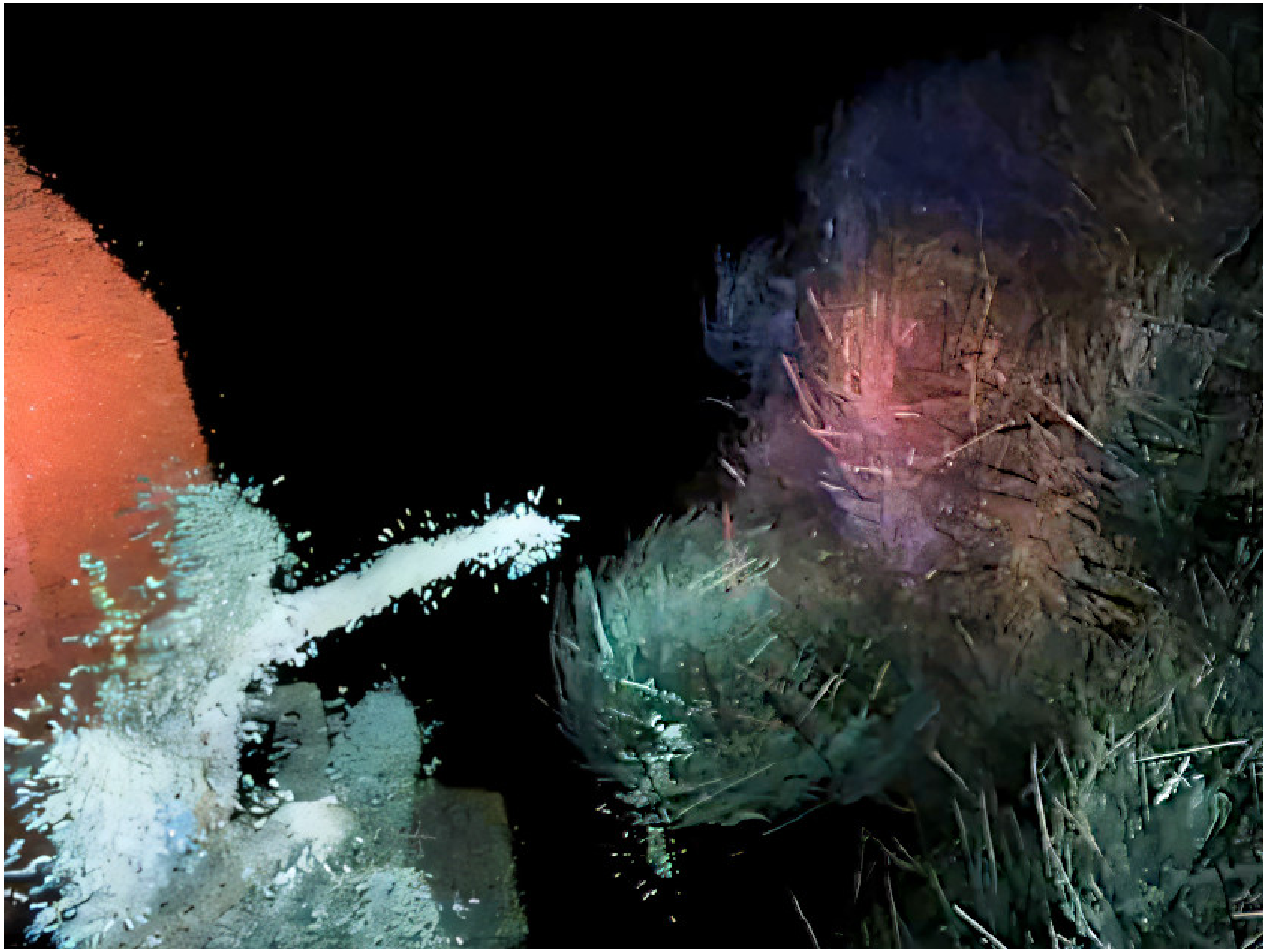


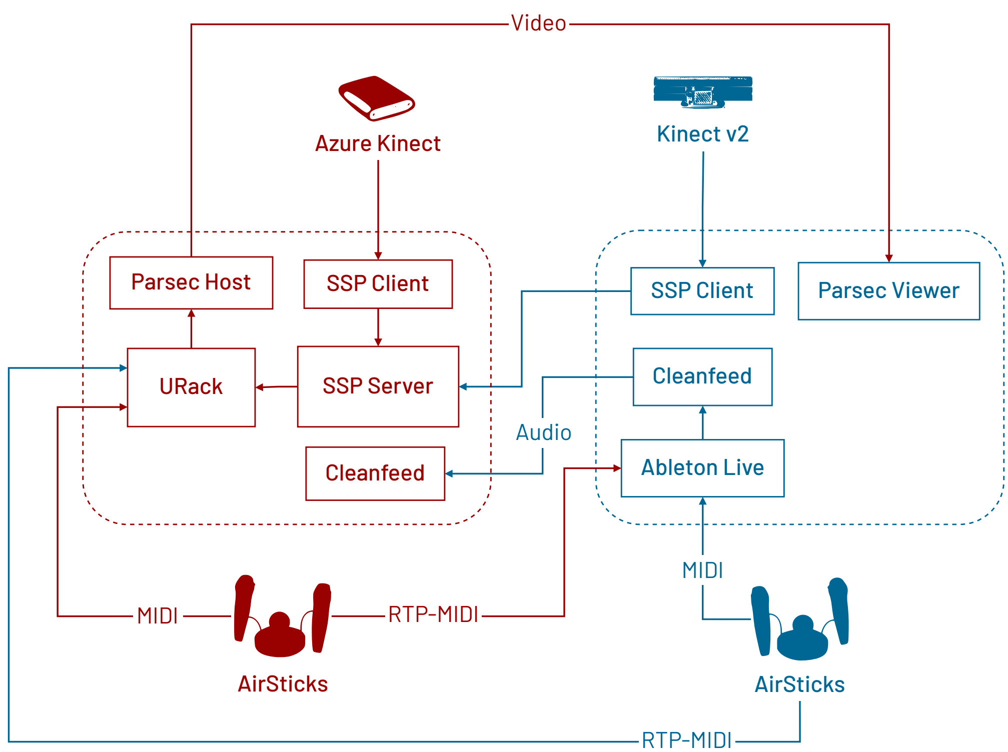
Figure 5.3 – AirStream *merges the rooms of two remote performers into a single audio-visual environment.*

5.2.2. System design

The systems developed for previous AirSticks productions all operate using a distributed network architecture, where data is shared between multiple computers tasked with distinct roles. In these performances, musical data such as OSC and MIDI, as well as image streams from depth cameras are transmitted over local area networks.

AirStream is no different in its distributed architecture. It contains several streams of data that must be shared between computers, though for this piece, these streams are transmitted over the internet. Figure 5.4 presents an overview of the distributed architecture of *AirStream*, highlighting components located in Sydney and Melbourne in red and blue respectively.

Figure 5.4. – *The system architecture for* AirStream *is distributed across two locations. Components physically located in Sydney and Melbourne are coloured red and blue respectively.*



AirStream utilises a *virtual private network* (or VPN) to facilitate the connection between the two remotely-located computers it uses. A VPN creates a private network over the infrastructure of public networks, allowing connected devices to communicate on the internet as if they were connected directly to the same local area network. A VPN allows us to use the same software and the same networked systems that we would use when on stage together — except over the internet — and without any need to think about running servers in the cloud or exposing our devices to the internet proper.

We utilised *ZeroTier*[1] to create our *AirStream* VPN, though there are alternative approaches. *WireGuard*[2] can be used, and is the de-facto standard modern VPN protocol. It can be configured itself, though services like ZeroTier exist to make the setup of WireGuard networks trivial. *Tailscale*[3] is an alternative which provides faster connections and a simpler setup than ZeroTier. Using a VPN for *AirStream* made orchestrating the communications over the internet incredibly straightforward as our connections could remain peer-to-peer, and we didn't need to spend time developing services that work in the cloud.

Sound

Ableton Live runs on the computer in Melbourne to generate music, and the set of AirSticks there manipulates sounds through a local MIDI signal. The MIDI signal produced by the AirSticks in Sydney is sent over the network using RTP-MIDI into the same Ableton Live session, where it is able to manipulate other aspects of the sound simultaneously.

---

1 https://web.archive.org/web/20230607164907/zerotier.com

2 https://web.archive.org/web/20230610160049/wireguard.com

3 https://web.archive.org/web/20230610174443/tailscale.com

The performer in Melbourne was able to listen to the music directly from their computer, and for myself to hear the Ableton session in Sydney we used *Cleanfeed*[4] to stream the audio. This solution provided a middling quality audio stream with considerable latency, resulting in a noticable lag between performing a gesture in Sydney and hearing a response in the returning audio. Some alternatives such as *Jacktrip*[5] or *Jamulus*[6] are worth investigating for future work to see if latency can be improved, though for *AirStream* we were able to accommodate for an audio lag in how we devised the performance.

The performer in Melbourne — who had the sound playing locally — created transient sounds by striking in the air with his controllers. The kind of rhythmic precision necessary to do the same in Sydney was not possible due to the audio lag, so my own gestures were mapped to timbral parameters for less rhythmic interactions.

Depth cameras

A depth camera at each location captures each participant's performance area. We used the cameras we each had access to. In Melbourne this was a Kinect v2, and in Sydney this was an Azure Kinect. The images from these sensors are published using a *Sensor Stream Pipe*[7] client running on each performer's machine.

Sensor Stream Pipe (SSP) is encodes depth and colour frames coming from 3D cameras and transmits them over a network. Is it built to send multiple video streams over a network in real-time. The compression provided by SSP reduces the bandwidth required for an RGB-D camera stream enough to allow transmission

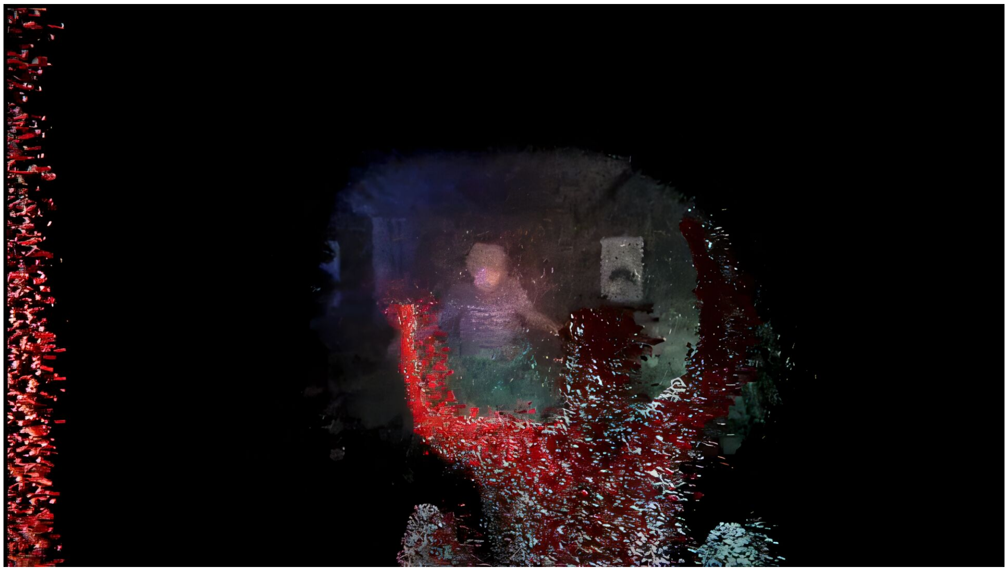
4 https://web.archive.org/web/20230601101303/cleanfeed.net

5 https://web.archive.org/web/20230316095313/jacktrip.github.io/jacktrip

6 https://web.archive.org/web/20230607144730/jamulus.io

7 https://web.archive.org/web/20230323035314/sensor-stream-pipe.moetsi.com

Figure 5.5. – *One performer has the ability to orient the view into the virtual environment.*



over the internet. SSP is suited for tasks where sensor data needs to be offloaded from the devices performing the capture and onto other devices for further processing.

The base SSP code makes some assumptions that are not suited for *AirStream* or live performance scenarios in general. In particular, it prioritises guaranteed delivery of each frame at the cost of increased stream latency. It also lacks support for the Kinect v2 camera that was available to us in Melbourne at the time. Due to these factors, custom changes to the SSP source code were made for *AirStream* in order to prioritise low latencies and to provide compatibility with the this work's hardware and visualisation pipeline.

The SSP streams coming from each location are received by the visualisation computer located in Sydney. Here, the streams are decoded into point-clouds and integrated into the unified virtual 3D scene. The graphics of this virtual environment are created and controlled using custom Unity-based visualisation software. The Unity instance in Sydney receives MIDI signals from both locations in a similar fashion to the Ableton instance in Melbourne.

Interactive video streaming

Similar to the audio, the graphics were rendered on a computer at one location. The Sydney location rendered the graphics and transmitted them to Melbourne as a video stream. To facilitate this, we used *Parsec*[8] — a high-quality, low-latency remote desktop service originally built to stream real-time gaming. Traditional video conferencing applications like Zoom were insufficient for our needs due to their inconsistent quality and poor compression artefacts. Similarly, basic remote desktop protocols like VNC and RDP didn't offer the performance levels required. Parsec was chosen due to its high-quality stream and impressive responsiveness.

The open-source alternative *Sunshine*[9] was also considered, though I found it does not handle unstable networks as well. It is able to offer higher quality video, though in situations of network instability, Sunshine can drop many consecutive frames or exhibit spikes in stream latency. In these cases Sunshine notifies the user that they aren't able to meet bandwidth requirements and suggests a lower bit-rate configuration. In contrast, Parsec proactively adjusts the video encoder's bit-rate to maintain a consistent frame rate when the internet connection is not able to meet throughput demands. This active adaptation to network conditions resulted in a consistent smooth motion at the cost of an occasionally more pixelated image, which was a trade-off justified for the liveness necessary in performance. In some cases though, the consistent high image quality offered by Sunshine might be a better fit, like if a high-quality internet connection can be guarunteed (generally not in Australia), or for applications that involve slowly moving or static graphics.

8 https://web.archive.org/web/20230206032124/parsec.app

9 https://web.archive.org/web/20230607101003/github.com/lizardbyte/sunshine

5.2.3. Virtual space as an instrument

AirStream presents a *playable* teleimmersive environment. This merged virtuality is not just a synthesis of each performer's space — it is an instrument itself. With two performers engaged, each player interacts with separate parameters of the same sonic and graphic parts, making the virtual space a single instrument for two players. Through their performance gestures, players simultaneously shape their reconstructed physicalities and conduct the tightly integrated soundscape.

The visual components of the virtual space are rendered as a dense particle system, illuminated in colourful and saturated area lighting.

The main auditory component is a bass synthesiser, which is versatile enough to work as a percussive sound, a main melody, or a drone depending on the state of its parameters.

The performer is afforded control over activating a note on this synthesiser. A strike to one of the eight spaces mapped out in front of his body allows him to play through a chosen scale. When a note is initiated, the particle system pulses outwards, and the colour of the scene's lighting is changed as the performer stretches this note and shifts the controllers around space.

By twisting the AirSticks, the performer controls a filter over the synthesiser as well as an orbital rotation of the particles that tracks against the filter's cutoff frequency.

For myself, I am afforded control over the level of a white-noise oscillator, and a ring-modulator. These are mapped to the 3D position in space of my set of AirSticks. As the white-noise rises, particles stretch into spikes. As the ring-modulation intensifies, the particles disperse into an abstract cloud.

This division of parameters that control the synthesiser was deliberately chosen. Since the performer in Melbourne had Ableton running locally, they could hear the audio without delay, so was chosen to be the one with rhythmic and melodic control over the sound. My own control affects only the timbre, which allows some tolerance in timing and somewhat bypasses the issue of audio latency.

The virtual camera represents a third presence in this environment. A setup similar to the 'drone' in *Computer Storm* is used, with zoom and orbit controls mapped to the analog sticks on the AirSticks. This allows for the orientation of the virtual camera while simultaneously playing the sonic and graphic controls.

AirStream reshapes the challenges of physical distance into a creative opportunity. It merges two geographically-disparate spaces into a unified, expressive instrument, showcasing the potential of teleimmersive environments. Beyond bridging the physical distance, we embody and manipulate the virtual realm itself, demonstrating how such spaces can serve as novel parts of the performance itself.

5.3. Layer

5.3.1. Displaced connection

Layer is an internet-connected dance production that embeds geographically-disparate performers in Australia and the Netherlands into mixed reality performance environments. Through the interplay of physical, virtual and fragmented realities, *Layer* probes the complexities of self-perception, isolation, and the emotional landscape of our changing world.

Over the course of the forty-minute production, the physical and virtual forms of the local and remote dancers are integrated into a range of mixed reality configurations where performer's physicalities are displaced.

Birthed through a shared experience of disconnection during isolation, *Layer* focuses heavily on *connection*. It narrates the team's collective journey around connection — its absence, its pursuit, and its realisation. It shows mixed reality used to elicit meaningful on-stage connections between remote performers.

Layer was presented over four nights — twice to a live audience in Sydney and twice again for an audience in Groningen. One Sydney performance was streamed to an audience online. Video documentation and a trailer are available below.

► **Video 5.2** - *A trailer for* Layer. *(03:00)*
Watch at phd.matth.cc/layer/trailer.

► **Video 5.3** - Layer *as presented at the Flying Nun in Sydney, October 2020. (36:14)*
Watch at phd.matth.cc/layer/sydney.

► **Video 5.4** - Layer *as presented at Grand Theatre Groningen, December 2020. A Q&A session with the audience team is shown at the end. (57:23)*
Watch at phd.matth.cc/layer/groningen.

Figure 5.6 – *A table placed in the performance space for* Layer *helps integrate the remote participant into the physical stage environment.*

Figure 5.7 – Layer *sees the trio of dancers inhabit multiple configurations of teleimmersive space.*

Figure 5.8 – *A fundamental theme of* Layer *is connection.*

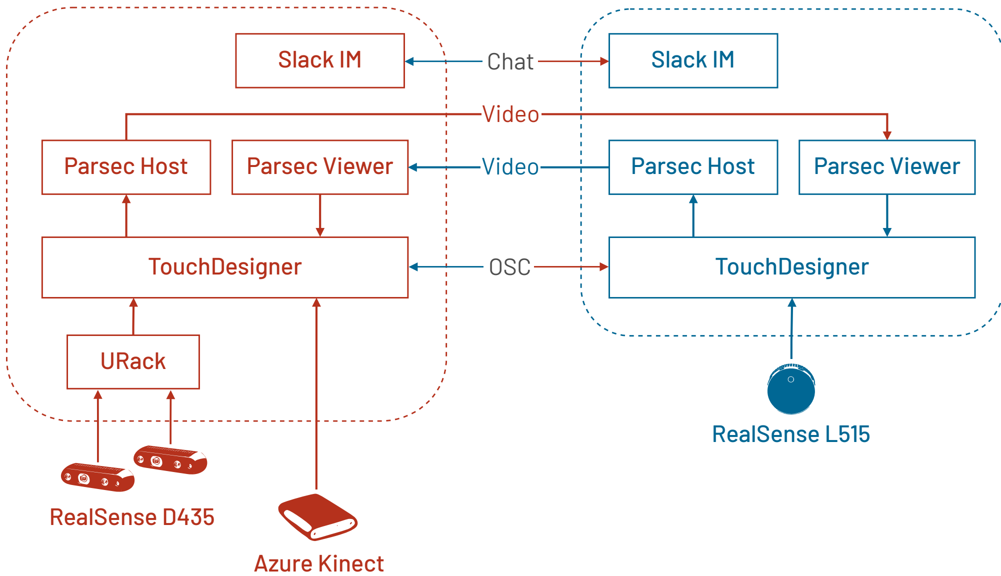
A massive team of people contributed to the development of *Layer*. This work was a joint-production between Box of Birds, Shelfish Productions and WERC. It was directed by David Clarkson in Australia and Veerle van Overloop in the Netherlands, and features dancers Olympia Kotopoulos, Cloé Fournier and Katina Olsen. The technical team in Sydney consisted of myself and Boris Bagattini, and in Groningen: Olav Huizer, Joachim Rümke and Jelle Valk.

The process of designing the system used in this performance was hugely collaborative. Each of the five experienced technical artists brought into the project their own idiosyncratic approaches to audio-visual system design. Through both purely local and remotely-connected development sessions, the technical team actively engaged with the dancers, choreographer, and directors. The technologies used and the design decisions made — as described in the following sections — are the direct result of this participatory approach. By considering each practitioner's current practice and knowledge, we were able to speedily create a system that integrated our workflows and attempted to support each stakeholder's creative goals.

5.3.2. System design

The network of devices and software used in *Layer* is illustrated in figure 5.9. Depth cameras capture performers in both locations as point-clouds, which are visualised in 3D, using TouchDesigner and Unity, before being transported to the remote location as streaming video. The remote video is composited with local visualisations and projected into the performance space.

Figure 5.9. – *The system architecture for* Layer *is distributed across two locations. Components physically located in Sydney and Groningen are coloured red and blue respectively.*



Connectivity

Layer builds upon the distributed infrastructure utilised in *AirStream*. Zerotier is again used for a virtual private network (VPN) between the Sydney and Groningen venues to share streams of data needed for the performance. Another unchanged component is the use of Parsec to stream remote-desktop video between the sites. These softwares mostly worked well, however we did face network connection difficulties during development.

The Sydney development of *Layer* took place in three venues. Two venues had consistent internet, but our most-used rehearsal space provided a lousy connection, and we were forced to use a 4G cellular modem. This introduced high latencies and lower available network throughput, which increased video lag and sometimes resulted in Parsec refusing to connect at all.

It is in the nature of this kind of experimental[10] performance development that teams may be required to move between different rehearsal and performance spaces. Because of this, it is incredibly important to have robust network infrastructure that can scale to work with poor quality connections, as internet quality is never guaranteed from venue-to-venue.

Even when quality internet is available, larger venues may implement restrictions on what connections are made available to network clients. It was not uncommon for the venues in both Groningen and Sydney to have a network firewall in place that prohibited Parsec from making a connection. A VPN is meant to remedy this, but we had trouble stopping Parsec from using the wider internet to first establish initial connectivity. When this kind of connection issue blocked us, we streamed desktop capture of our computers over the video-conference application Zoom[11] instead, which presented a poor quality stream unsuitable for final presentation, but tolerable for use when developing choreographic ideas. Since video-chat applications like Zoom send their data over TCP using standard HTTPS ports, they are less likely to be blocked on a foreign network. High-quality remote desktop such as Parsec or Sunshine send UDP traffic over uncommon ports, which makes them more likely to be blocked if a venue has an IT team that values internet security. For some venues we needed to request for Parsec's ports to be unblocked by IT.

Where *AirStream* (5.2) combines incoming data together into a single virtual environment rendered only once, *Layer* requires that two separate visualisations be rendered in order to reproject the remote performers into the opposite environments simultaneously. Visualisations of the performers and a series of virtual environments are created using a combination of TouchDesigner and Unity. Computers in both locations send their visualisations into a TouchDesigner instance

---

[10] read: underfunded

[11] https://web.archive.org/web/20201018014820/zoom.us

that acts as a 'video mixer', compositing local visualisation streams onto the incoming remote streams from Parsec. These two merged visualisations constructed on each side are projected onto their respective stages. The end result is two (sometimes dramatically) different presentations between locations that contain the same live performances, but are influenced by the aesthetics decided upon by each team.

Data is not sent directly from the depth cameras themselves over the virtual private network, instead the point-clouds are placed in a 3D scene that is rendered locally and transmitted over video. While this means a shared 3D virtual environment is not generated directly as it is for *AirStream*, 3D control over the remote visualisations is still afforded to the receiving team. The remote desktop connection allows one team to take control of their partner's computers in order to orient remote virtual cameras. The remote scene can therefore be lined up according to the requirements of the local virtual and physical environments.

For some scenes, just the remote performer is projected onto the stage. This is shown in figure 5.10), which effectively works to teleport the remote dancer's presence to the in-situ performers by placing her virtuality on top of a physical table. In this case, the Sydney tech team takes control of the Groningen visualisation to transform, scale, orient and otherwise align the 3D representation of the remote dancer with the physicality of the local performers and the table.

In other scenes, each location's visualisation is integrated into a teleimmersive rendering (as shown in figure 5.7) that is projected at the rear of the stage. This projection acts as a viewport into a virtual location that is separate from both physical stages.

Music is not streamed directly from one location to another. We found that it was not worth straining the bandwidth of our already inconsistent network connection. Parsec itself can send desktop audio, though we in unreliable conditions this was

Figure 5.10. – *A remote performer dances with two performers in-situ on* Layer*'s mixed reality stage.*

the first part of the stream to deteriorate. None of the music presented in this production is performed live, so it is not necessary that audio be streamed in real-time. The only requirement is for the playback of audio to be synchronised. A playlist of the prerecorded music as wav files is loaded in both locations inside a TouchDesigner patch. OSC messages sent between each location facilitate the synchronised playback of the files on each side.

In addition to the video and audio communications visible to the performers and the audiences, a vital component of the connectivity between our remote performance environments also worth mentioning is the use of instant messaging. The synthesis of our remote environments required the management of several networked computers transmitting many streams of data. Orchestrating these streams and their operation smoothly is crucial for a successful performance, so while the dancers can communicate through their choreography, projected as video, the tech teams across locations are constantly communicating through instant messaging to coordinate scene changes, trigger audio-visual cues, orient virtual cameras, and troubleshoot the systems in real-time. We used *Slack*, although any private chat-room is equivalent.

Depth cameras

Layer used a range of different depth cameras to capture the performers in 3D. The Sydney team utilised two Intel RealSense D435 structured-light cameras, one Intel RealSense L515 time-of-flight camera, and one Microsoft Azure Kinect time-of-flight camera, while the Groningen team used a single Intel RealSense L515 time-of-flight camera.

Each camera has its own qualities, which are described more broadly in section 6.2.2 on the practical considerations of 3D camera selection. However in relation to their specific use with Layer, these cameras had direct impact on choreographic choices made during development. One scene in particular (shown in figure 5.7) was crafted around the distortions inherent in the quality of the cameras.

During this scene, the Sydney side use the two D435 structured-light cameras on either side of the stage, with performers sat directly in front. This camera produces a 'waviness' in the depth image it creates. While this distortion makes depth maps produced by the camera unusable for applications that require accuracy, these cameras were useful in creating a deformed, grotesque image that is used to convey a sense of discomfort in how the self is captured and presented through technology. One performer commented on how "ugly" this particular camera made her face look, and the choreography was crafted to amplify this ugliness by having the performers tear at their faces.

The Groningen team made use of the L515 time-of-flight camera, which is poor at capturing human hair (among other specific materials). The L515 is only able to consistently detect hair if it falls on-top of other parts of the subject's body — such as the shoulders, chest or back — with any long hair that extends out from the body's silhouette usually rendered invisible. This fault of the camera technology again influenced the choreography of this scene, leading to all three performers alternating between extending their hair out from their heads and wrapping it around their bodies.

5.3.3. Modes of Interaction

As we experimented with different ways of having our on-stage and remote dancers share their performance space, a primary focus of the development of *Layer* was the investigation of a number of different teleimmersive interaction modes.

Although our initial goal was to create a piece that could be experienced by two audiences at once (one in Sydney and one in Groningen), due to differences in time-zone and venue access, we did not end up presenting *Layer* to both audiences simultaneously. This meant that during each performance, remote performers were essentially "supporting" the location that did have the audience. Figure 5.11 shows the Sydney team playing the role of support during a live Groningen performance. While not our initial intention, this allowed both teams to explore the interaction modes more suitable for the venue and available resources, making *Layer* essentially two different performances.

Notably in the presentations for Sydney, we included scenes focused around the exploration of *telepresence*, while the performances for Groningen placed a greater focus on *teleimmersion*. When local and remote performers in Groningen are presented together, it is achieved by transporting all performers virtually into

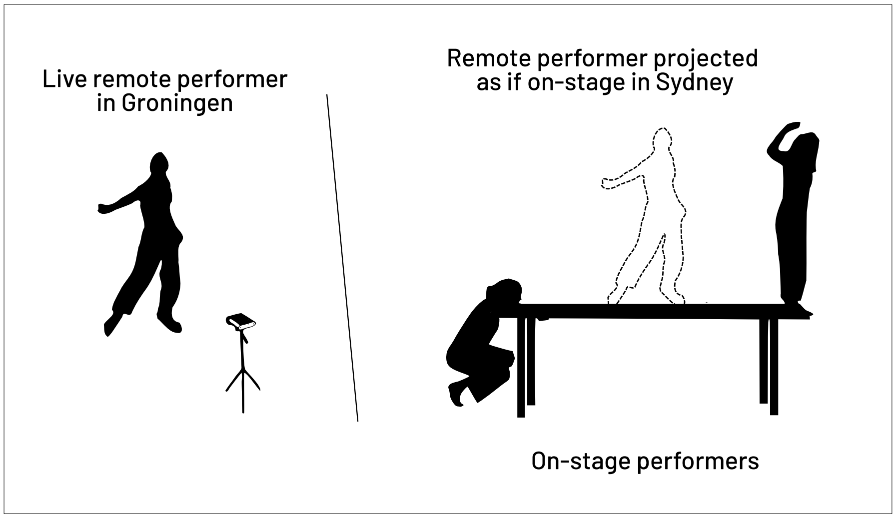
Figure 5.11. – *The Sydney performers of* Layer *are captured and streamed to the live audience in Groningen. The monitor on the left displays the live performance happening in the Netherlands, and the engergy drink illustrates the effect of working across time-zones.*

shared virtual spaces that are projected detached from the local performance environment. The Sydney presentation however, also explores how a remote performer's presence can be placed into the physical space of the stage among local performers.

The opening scene showcases the attempt at *telepresent* performance, where the audience in Sydney see the remote performer projected life-size on top of a physical table (shown in figure 5.10). The choreography in this scene aims to create a trio as if the remote dancer was also in the Sydney performance space.

Though the remote dancer herself was not dancing on top of a table in Groningen, having her projected above the table in Sydney added to the illusion she was sharing the stage. The engagement with the physical set-piece by both local and remote dancers was extremely effective in grounding a presence in the performance area for all participants. The scene also features moments where

Figure 5.12. – *The table scene in Sydney's version of* Layer *placed the remote dancer into the physical performance environment.*



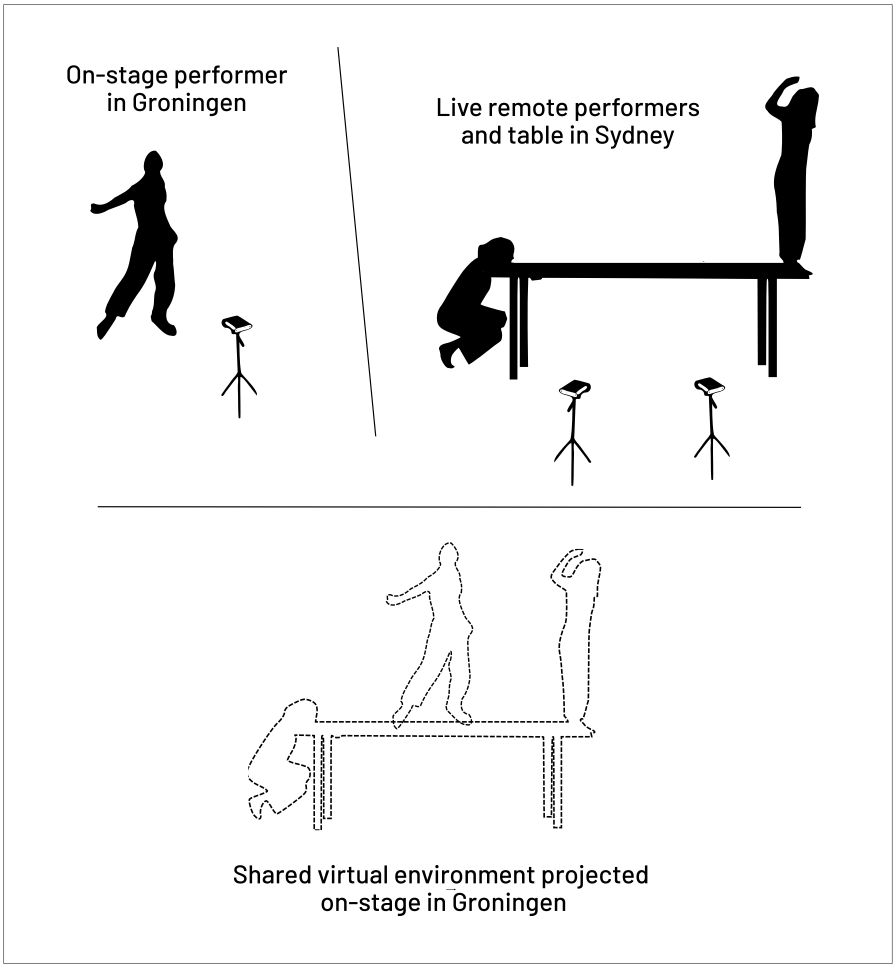
our live performers move *into* the projection of the remote dancer, conversely attempting to showcase her genuine distance from the audience, as well as her digital representation's ephemerality.

The telepresent configuration as presented in Sydney is illustrated in figure 5.12. For the Groningen showing of this scene, the performers in Sydney and the table they dance on are captured and placed into a virtual environment that the performer in Groningen is also placed into.The fully teleimmersive configuration of this scene as presented in Groningen is illustrated in figure 5.13.

5.3.4. Geographically-disparate performance development

With our partners operating eight hours behind our time zone, we could not put in full work days together like an entirely local team would have during a conventional

Figure 5.13. – *The table scene in Groningen's version of* Layer *placed all dancers into a shared virtuality that was then projected beside the performance environment.*



development period. Instead, a large amount of our collaboration was asynchronous. We found ourselves experimenting with technology and choreography in our daytime, while our partners were asleep. This would be reciprocated by the Dutch team, who would undertake their own explorations while we were asleep.

In addition to using *Slack* during the show for real-time communication, we relied heavily on the group messaging platform to share our thoughts, explorations and discoveries. Each morning we would wake to a collection of ideas and experiments the Dutch had posted. We would then spend the day expanding on their concepts

Figure 5.14. – *The remote performers explore the table scene together.*

and incorporating their ideas into our work, layering our own experiments on top, then posting the results back to them. This reactionary approach turned into a rhythmic cycle of iteration between our two groups.

Once we had some choreography set, our days of development included local dancers practising alongside recordings of the remote half of the performance projected onto the wall (as shown in figure 5.14).

A few times a week we scheduled live online rehearsals together, to bring together what we had done offline and explore it cooperatively. These joint sessions were when a real energy could be felt.

> When we do the show ... or we do a rehearsal [by ourselves], it feels ... very, very flat. But as soon as ... the three dancer's are together, there's a magic that happens ... they're 12,000 kilometres apart or whatever it is ... but you feel the presence of the third person. — *Director*

We spent significant amounts of our time in these online rehearsal sessions troubleshooting and refining the technological aspects of the work. We dealt with connectivity issues and were forced to spend this connected time optimising our network and coordinating the different video streams coming and going from multiple computers. The technological problem-solving that we needed to focus on to just get a collaborative performance working limited our ability to simply be creative together.

> In terms of getting all this information around, and distributed, and connected … just [on] our own computer systems, and also between *our* computer systems and *their* computer systems … Because of the really limited time that we got to … connect together, a lot of the time was spent troubleshooting technical stuff … so there wasn't really a lot of direct creative time that we got to spend. — Digital Artist

An ideal system for a development like this could be based around the ability for teams to easily join an online space that combines their performance environment with other connected locations. There is a need for a simpler system for performance-focused telepresence to exist that handles connectivity, communications and streaming seamlessly so that the team in the room can focus on making artistic decisions about the work, rather than wrestling with the technology. Much of the motivation for the pointcaster system described in section 6.2 comes from a desire to make this practice easier.

5.3.5. Presence

Layer was successful in imparting a sense of presence onto remote performers projected into the space. The dancers who were physically present felt a strong connection with their remote partners, as if they were sharing a physical space.

Figure 5.15. – *The remote performer takes a break from rehearsal — relaxing underneath a tree inside the locally-rendered virtual environment.*

> It feels like she's in the room even though she's not. It's really weird but it feels like she's present. — Performer

The performers developed a natural sense of awareness of their partner's presence and became accustomed to imagining they were in the performance area together. We had screens positioned so that the performers could guage their position in relation to their virtual partners, though through many of the scenes, these monitors went unused.

> I didn't have to look at the monitors at all ... I could figure out where Olympia was, so then it was just about adjusting my gaze ... this was very intuitive. — Performer

The genuine sense of connection to the remote participants felt by the dancers was described as a transmission of *energy* into the distant space. *Layer* was a dance piece that in the end became less focused on synchronised movement and more centred around connecting these women's remote physicalities.

> I think the intuitive part was the *energy*, because actually in the box I don't "dance" much — it's more angles of the body ... How you send focus into this blank space that then gets transferred to them. That's I think, the intuitive part ... I can do a movement and it looks nice on the screen, but it can really easily be disconnected ... so it's just more [connecting] energy than anything else. — Performer

The dancers commented that the latency present in the video streams wasn't a major hindrance to their practice. They generally had no issue 'locking in' with their remote partners. The choreography in this performance does not rely on strict musical rhythm like other less-abstract genres of movement might, which undoubtedly made the internet delay less crucial.

While it didn't make performing the choreography all that much more challenging, one dancer stated the mere existence of latency amplified the feeling of distance between her and her dance-partner.

> It feels further away when there's a delay with the internet... — Performer

Even if the delay didn't affect the performance of choreography with this particular style of movement, it does inhibit feelings of connectedness. To further support the sense of presence described by the performers and properly stimulate feelings of co-location, reducing latency needs to be a key technical focus in future work.

5.3.6. My House Your House

My House Your House — or *Mi Casa Tu Casa* — was a follow-up project to *Layer*, continuing an exploration of teleimmersive and telepresent performance. It uses essentially the same system design, which is why it is not given its own section in this chapter, but it is still worth mentioning.

This work again aimed to transcend geographical boundaries, streaming performers live from a dance studio in Tijuana, Mexico onto the stage at a techno music event in Sydney, Australia — for which a highlight clip can be seen in video 5.5. *My House Your House* focused on exploring the teleimmersive space created when fusing the 3D camera feeds of both locations into one.

A Kinect v2 camera in Tijuana captures three performers, and an Azure Kinect in Sydney captures another three. Both location's digital scenes are integrated for display in Sydney, although this time the visualisation is projected onto a scrim at the front of the stage (as pictured in figures 5.16 and 5.17).

► **Video 5.5** - *A highlight video of the electronic music event that featured telepresent dancers streamed in from Mexico.*
Watch at phd.matth.cc/micasa/recap.

My House Your House featured the Lux Boreal Dance Company joining Box of Birds, with performance by Phillipa Keogh, Josh Freedman, Madeleine Backen, Ángel Arámbula, Raúl Osuna and Henry Torres Blanco, choreography by Cloé Fournier, music by Nathan Moas, and technical production by myself, Boris Bagattini and Mario D Alvarado.

With the remote performers situated on the scrim between the local performers and the audience, our local performers had an easier time maintaining a connection with both their remote dance partners as well as the crowd. The dancers in *Layer* reported strong feelings of presence from their remote partners, and

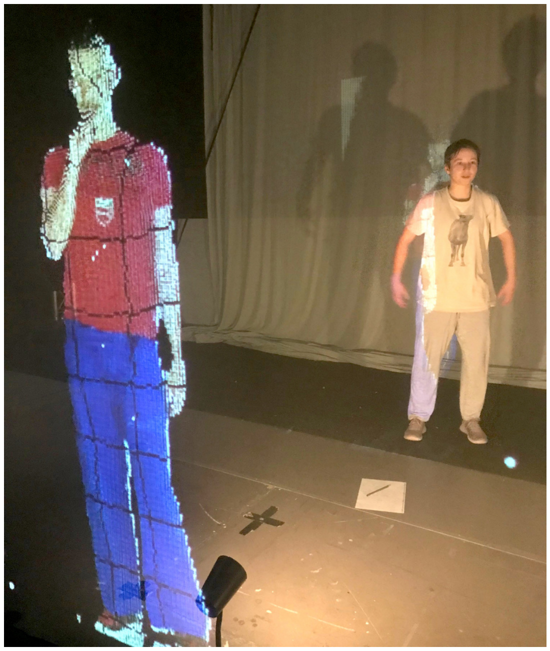
Figure 5.16. – *A dancer in Sydney is projected onto the plane of virtuality at the front of the stage, enabling a geographically-disparate duet with the performer in Tijuana.*

could intuitively localise this virtual presence in physical space when they were not looking directly at the projection surface. However, parts of the choreography for *My House Your House* were much more improvisational than *Layer*, necessitating more constant visual communication between dance partners. As a result, the use of the scrim was essential in facilitating a reactionary performance, and supported a more fluid interaction.

As discussed with *Layer* an obvious latency is introduced when making connections over the internet from locations this remote. During our sessions linking Sydney and Tijuana, it was not unusual for the video feed of the remote dancers to be delayed by up to a second.

However, the choreographer used a fascinating dance exercise called *flocking* to overcome this delay and promote a synchronicisty between all the performers involved. Flocking, inspired by the motion of bird flocks and schools of fish, requires each dancer to follow simple rules in response to the relative position and

Figure 5.17. – *The scrim supported visual communication with performance partners and the audience.*



movement of their fellow dancers. The exercise balances cohesion and repulsion, and the role of leading movement in the group seamlessly fluctuates as participants navigate gesture through the shared space (Leonard et al., 2012).

The six dancers of both locations were integrated into the virtual scene and projected in front of the performance area, where this latent connection of group dynamics was used in order to produce tight synchronisation of movement between participants. The continuous and fluid adaptation of flocking allowed the group to absorb any latency in the video feed and incorporate it into their collective movement. This technique was used as an exercise to encourage connection between the remote groups, and was used at the start of the performance to allow the dancers to 'lock-in' with each other amidst the lag of the internet. The latency became a part of the performance, rather than a hindrance.

5.4. Critical Path

5.4.1. Displaced physicality

Critical Path shifts performers in space and time by capturing their physical forms before reconstruction and displacement within augmented reality glasses. This work explores connection, spirituality, and the boundaries between the physical and digital. It integrates live performance with 3D telepresence, allowing performers to engage with one another remotely in a space shared by physical and virtual participants.

Audience members, equipped with the AR glasses, roam freely through the performance space among copied and transported 3D streams of the dancers captured in real-time. Mirrors placed around the performance area further transpose the presence of performers and spectators.

Documentation was captured using cameras on the AR glasses. The quality is not high, though it gives a fair depiction of a participant's experience. The first two performance videos below and are filmed from a static position. A third video shows a first-person AR view during development to illustrate the dimensionality of the experience.

► **Video 5.6** - *The* spirit duet *scene sees two virtual performers dance together before one takes physical form. (02:28)*
Watch at phd.matth.cc/critical-path/spirit-duet.

► **Video 5.7** - *The* virtual sculpture *is created by capturing and layering moments of a dancer through time. (04:36)*
Watch at phd.matth.cc/critical-path/sculpture.

► **Video 5.8** - *First-person footage of the* virtual sculpture *scene in development shows the agency participants have over their viewing perspective. (01:37)*
Watch at phd.matth.cc/critical-path/sculpture-dev.

Figure 5.18 – *AR glasses combined with mirrors bring about multiple projections of the performer's presence throughout the physical environment.*

Figure 5.19 – Critical Path's spirit duet *teleports remote dancers onto the stage using augmented reality.*

Figure 5.20 – *The virtual sculpture scene in* Critical Path *sees moments in time frozen and layered as the piece progresses.*

5.4.2. System design

Critical Path was conceived on the idea of harnessing augmented reality glasses to bring the telepresent performance explored in *Layer* into three dimensions. For this initial exploration of performer displacement, the 'remote' performers that are translated onto the stage are within the same room, on a local network. The ultimate goal however was in constructing a system that is suitable for global telepresence.

The system designed during the development of *Critical Path* is called *pointcaster.* As the final presented system design in this thesis, it represents a culmination of the knowledge gained through development of many mixed reality performances. *pointcaster* is therefore discussed in greater detail with respect to its larger context in chapter 6. For now, I will provide a less-detailed overview of the system in the context of *Critical Path.*

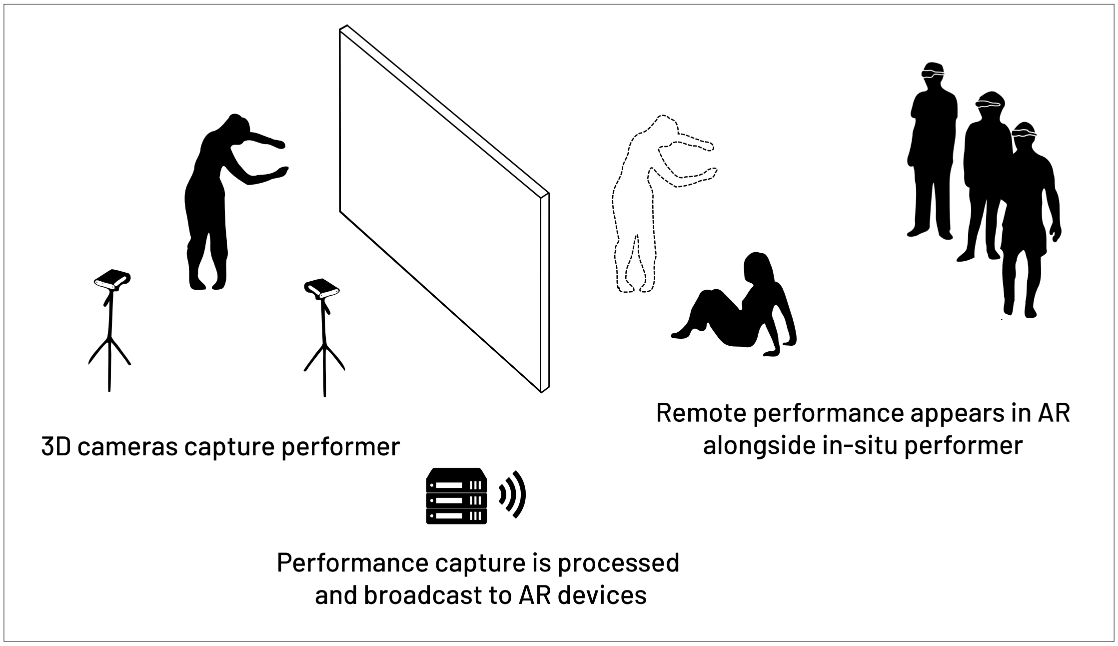
This work uses three Azure Kinect depth cameras to capture performers as point-clouds. At times, cameras are pointed at multiple performers, and at other times, multiple cameras capture a more-complete rendering of a single performer. The cameras are connected to the pointcaster server, which constructs a virtual scene from the incoming frames and broadcasts this to the augmented reality glasses over WiFi.

The glasses are six pairs of XReal Light[12]. Each set is driven by an Android smartphone, running a Unity-based application that receives the broadcast from the pointcaster server and renders the live virtual scene in the headset.

A high-level overview of how telepresence is implemented in *Critical Path* can be seen in figure 5.21.

[12] XReal, formerly known as NReal. https://web.archive.org/web/20220502234950/nreal.ai/light

Figure 5.21. – *High-level overview of telepresence in Critical Path.*



5.4.3. Increased physicality

One of the most notable qualities of the system of AR telepresent interaction utilised in *Critical Path* was a sense of *physicality* unmatched in my prior mixed reality works presented using projection. An increased perception of physicality was described prominently in the interviews with collaborators, was noted by audience-members after experiencing the piece, and can be observed as a clear focus in many of the behind-the-scenes videos captured during development.

> With the glasses, there is this illusion that your body is actually involved.
> — Choreographer

Compared to using scrims for telepresent projections like our prior works, the use of AR headsets provided a greatly amplified feeling of physicality for audience-members.

> [Projection] lacks a physical engagement from the audience ... they couldn't really feel it, I didn't think. When I was with Linda, trying to virtually touch each other, you feel it ... there is a visceral response ... and it's not just novelty. — Director

This sense of physicality allows participants to feel more connected with the telepresent bodies in comparison to when images are projected onto a flat surface. One performer described a significant difference in the *connection* and the *presence* felt when moving alongside telepresent images using the AR headsets compared to 2D projections.

> With the [AR] glasses, I could feel the presence through the layers of the skin. There was a tactile sense. Where with [the projection] it was more my brain was connected ... I knew my movement of course, and we had a connection ... but the connection is not as strong as with the glasses. It really feels like it's another body in the space ... [In Layer], the presence was more the image. It was a creative connection with an image. — Performer

One curious point to note was how the feeling of presence and connectedness with the reprojected images could be retained when participants were to remove the headsets entirely. This was acknowledged by both those involved in creating the work, as well as the audience members.

Due to the physical constraints of the devices (notably the cable between the headset and the computing device, as well as the bulk of the computing device itself) *Critical Path* was presented without the performers wearing headsets themselves, with only the audience using the devices to view the augmented scene. This freed our performers to utilise their full range of movement. The choreography was developed by wearing the headsets however, and the performers were able to utilise the one-to-one dimensionality of the AR scene to position themselves into known real-world positions. They were then required to recreate precise

Figure 5.22. – *The performer felt she could sense the disembodied hands floating above.*

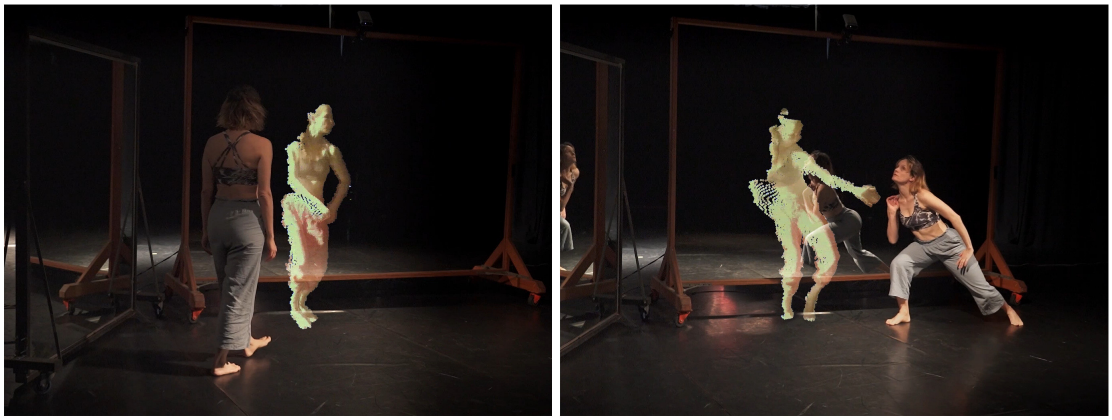
movements so that they could achieve the choreography devised during headset use. Once this choreography had been set, the feeling of presence was still felt by the performers when the headsets were removed. One performer noted this when speaking on a scene with another performer making hand gestures that are reprojected into the space above her as she was performing floor-work (shown in figure 5.22).

> I could really sense the hands up there, and I felt I was really playing with the hands... — Performer

Likewise when reflecting on a scene that reprojected the second performer's body into the performance space to create a hybrid physical-virtual duet (as seen in figure 5.23), one performer again noted this illusory feeling of presence.

> ... she was behind a mirror. I couldn't see her but I could sense her. — Performer

Figure 5.23. – *A duet scene in* Critical Path *presents the two performers first as fully-virtual. After one performer physically enters the stage, we see contrasting interaction between one physical and one virtual performer.*



Even more curiously, during a session where we had too many audience members show up to facilitate with the number of headsets on-hand, multiple participants who were watching the work without a headset noted that they still felt as though another body was in the space during the hybrid duet scene. The presence described here reinforces the success of the duet. It was not just two solos performed in separate locations, with one digitally overlaid through the headset, but it really was a successful exchange between two dancers connected via the technology.

The increased physicality and presence that participants experience through presenting telepresent performance using AR headsets tricks the brain into a feeling of genuine tactility. Seeing either oneself or another person in AR, being able to navigate around the reprojection in 3D space, as well as the opportunity to situate oneself inside the "personal space" of the reprojected body leads to a unique kinaesthetic experience. An illusory sensation of touch was described by all of those involved, and nearly every participant attempted to reach out and touch the images presented in the headset.

> In this first session we were really obsessed with trying to touch each other ... because it was so phenomenally strange. — Director

> There's a different phenomenology happening ... You become like a child. You just want to touch it, everything, and you know you can't touch it, but somehow your brain doesn't quite comprehend it. — Performer

The kinaesthesia described by participants is not an experience totally unique to augmented reality headsets. The director of *Critical Path* previously worked on an installation for a CAVE-like environment utilising stereoscopic 3D projectors, described in a paper by A. Bluff and Johnston (2017). The authors describe the "phantom temperature and touch effects" experienced by participants as they become surrounded by virtual fire, and as butterflies appear to land on outstretched arms. However, when prompted to compare this phenomenon between the stereoscopic CAVE environment and the technology utilised for *Critical Path*, the director admitted the kinaesthetic effect was enhanced this time around — particularly due to the inclusion of telepresent bodies within the experience.

> That moment where the two virtual bodies touch, or try to touch and that kinaesthetic mind-fuck, I guess, of how you interact with a virtual person like that — that was the thing we just kept on coming back to over and over again ... because it does mess around with the kinaesthetic proprioceptors ... the difference between that and doing large scale interactive projection on a 3D surface is there was a kinaesthetic kick that came in ... we had a lot of excitement when we were starting to work in ... 360 3D projection ... you hold out your hand and the birds fly to you or whatever, but it still was not the level of kinaesthetic engagement that you get when it's an engagement with a virtual person. There's something new about that. — Director

Figure 5.24. – *Footage captured in our first development session shows performers, replicated and transported across the performance space, moving through each other's bodies.*

A significant portion of the first-person headset video captured during the initial development session of *Critical Path* features participants simply touching the reprojected images of one another (shown in 5.24). Performers experimented with moving their real and virtual bodies through the reprojected versions of each other and themselves.

The bizarre kinaesthetic experience that this combination of technology delivers evokes a strong inclination in participants to interact with the work by attempting to reach out and touch the images. As such, presenting a work like this inherently stimulates audience participation, which is a feature creators can take great advantage of.

Combined with the kinaesthetic effect, the 3D nature of the technology and the dimensionality of the perceived images encourages users to navigate and explore the performance space to experience the AR headset's full effect. This positions AR headsets as ideal for use in non-traditional performance venues. While presenting a work in a proscenium-arch stage environment does showcase some of the increased dimensionality provided by the use of AR headsets, interaction can be pushed further when participants are able to explore freely and navigate themselves around the projected images.

> Manipulating the gaze ... we do that in a live show ... manipulate where you want the focus to be, but it's just the eyes that are working. I think with the glasses there's a potential [to manipulate] where their body is in space instead of having [the audience] static. If you have them static then suddenly you go back into that "fourth wall" where the audience is just an audience. — Choreographer

The participatory nature of the technology can be empowering for an audience, but it also forces creators to anticipate how an audience could interact throughout the work. Participants are inherently encouraged by the technology to explore, which requires creators to think about how composition appears from an infinite number of vantage points.

> When you make choreography, you just make what you like, but with the glasses the whole time I was really thinking about the audience. I realised it was all about the audience ... more than a show where you just sit and watch, where pretty much everyone has a similar experience. I think with the glasses there is that "choose your own adventure" a bit more. Which is empowering for the audience. — Choreographer

This point became particularly salient when observing the effect different camera configurations had when capturing bodies to be reprojected into the headsets. With single-camera setups where depth images were only being picked up by a single sensor, some vantage points lacked the dimensionality that can be provided by the technology when fully 3D images are captured. The output created by a single camera produces what's known as a *2.5D* image. When visualised in a point-cloud, a *2.5D* image renders points that appear to emerge from a single origin plane.

This effect is most visible in figures 5.23 and 5.25. Notice the lack of fill in the backside of the body — this is due to the single depth camera being positioned in front of the performer. In other scenes (such as shown in 5.20), captures were filled to be fully 3D by adding more cameras and aligning them to each other in software. The 2.5D projections created weak imagery from particular vantage points, so our choreographer focused on trying to hit vantage point 'hot-spots' that would amplify the visualisation's dimensionality. She became attentive to the angles the performance was successful when viewed at, and was able to tease out more striking graphics by placing the physical and virtual bodies through a range of angles — in her words, she needed to harness the 'diagonals'.

> You really needed to have [the audience] in a particular point otherwise it becomes completely flat ... that's where we discovered that the diagonals were so important ... when you work in 2D projection, diagonal is not as good, but in a 3D context, [it's] so important ... Flat from the front works too, but I think there's very good hot-spots in the diagonals.
> — Choreographer

The angular approach to composition in this work was a consequence of exploring the 2.5D images created with a single camera, but it was a strong premise that continued when we transitioned to multi-camera, fully-3D capture setups. Using all possible angles for movements and positioning, generally taking a fully *radial* approach to composition is a good way to harness and showcase the *dimensionality* provided by the stereoscopic AR glasses. It also lends favour towards the increased affordances participants have with choosing their viewing orientation. It is substantially different to creating choreography to project on a flat surface (as with *Layer*, section 5.3), where the differences between small angles go unnoticed, and choreography in that context is all about either parallel or orthogonal movements against the plane of the image itself.

> There's this idea when it's 2D and it's flat, it's all linear so the choreography becomes linear. With the glasses it becomes circular — I like to imagine it like a little globe. — Choreographer

The radial composition approach is illustrated well in figure 5.20, which presents a scene where the digital reconstruction of the performer has been frozen in moments over time. A 3D sculpture is created from these moments with the performer suspended in different angular orientations, forming interest when viewed from all perspectives. The dancer, hidden off-stage at the start of the scene, joins the virtual effigy and physically moves through the instances of her own past.

5.4.4. Challenges of AR telepresence

Developing and presenting a telepresent performance work through the medium of AR headsets brought about many challenges. Some of the difficulties presented to the team were inherent in the medium, wherein characteristics of the technology required us to reason about composition in new ways. However, some of the challenges presented to us were simply due to the immature nature of the technologies used. The expectations of what could or should have been possible with the technology were at times not met, and the technological limitations forced the team to make undesired compositional concessions.

> With the [AR glasses] I feel like you need like three months to make twenty minutes [of choreography] ... it's a slower process because our brain is not yet wired to think that way. — Choreographer

The limitations of the technology made the development process difficult, but as many creators know, limitations in creative processes give birth to innovation and growth as artists.

Figure 5.25. – *The choreographer and performer both wore headsets while initially creating the choreography for Critical Path.*

> I knew it was going to be tricky ... but it really expanded my mind ... I think now if I go back into a normal — without glasses — choreographic process, I consider the space differently. Looking from all the angles ... it developed me as a choreographer. — Choreographer

Mixed reality performances require a focus towards realising composition that occurs in two separate but overlapping environments — the digital environment and the physical. Ideally these spaces combine to become a single digitally augmented performance environment, and this fusion does occur to an extent — particularly when participating only as an audience-member. But when developing *Critical Path*, the team found it hard to work as though both spaces were one.

One theme that revealed itself repeatedly in interviews with participants as well as in behind-the-scenes video footage was participant's difficulty in reasoning about the spatial orientation of telepresent bodies inside the reprojected space.

> It was a real gymnastic of my brain to figure out ... my right from my left when I was in the other space ... in the glasses space. — Performer

> Because of the orientation of the cameras, we had very different directions. In a way, our north was not the same ... I was guiding her through what I was seeing, but she didn't know what I was seeing ... There's a lot to learn about how you communicate with the dancer. — Choreographer

A performer must take direction based on inhabiting an alternate 3D space — the virtual space they are reprojected into digitally — which is a complex exercise. Establishing a clear vocabulary of spatial orientation is essential when producing a work that incorporates these technologies. A lot of time could have been saved during development if there had been effective communication between our choreographer and performer.

> It's a work of imagination, and I should have imagined myself projected into her direction ... Sometimes [choreographers] work with a numbering of the space ... instead of telling her "go right", "go left", I should have said "go towards three", or "go towards four". — Choreographer

For *Critical Path*, we ended up utilising a huge number of tape-marks on the floor to help performers orient themselves in the virtual space. The precision of performer's movements became a major focus of the choreography — certainly more than in our previous telepresent developments that projected bodies rendered in 2D.

> With the 2D you can be a bit less precise and it still works ... it's not about the subtleties with 2D. It's more about the general look. With 3D [AR] glasses it's all about the subtleties because you can make it really fit well. With 2D the audience won't see the difference. — Choreographer

The extra attention required to precision of movement ended up frustrating our choreographer. This effort was noticeable in the behind-the-scenes video and was also noted during the interview.

> I know that behind the mirror, I was a bit annoying with that because I knew where I wanted her exactly to give that illusion, and the millimetre forward it was just not the same effect. — Choreographer

5.4.5. Limitations of precision

While the previously mentioned challenges we faced during development were inherent in the choice to use AR headsets as our presentation medium, some limitations were primarily due to the current state of the technology. Many of the following problems I expect will be remedied as AR headsets become more mature. The state-of-the-art is closer to solving some of these problems than others, and it may be true that some issues will be less apparent on higher-end hardware than what we used for *Critical Path*. However, creators should take the following issues into consideration when producing similar works. More detailed discussion on the constraints imposed by current-generation hardware appears in chapter 6.

As previously noted, the nature of the three-dimensional presentation medium required a high degree of precision from the performers in order to properly satisfy our choreographer's creative vision. At the beginning of development, we discovered some strong images where the virtual reprojection of one performer appeared to be touching a mirror inside the physical environment with her hand. But quickly we began to understand that any precise images we produced were ephemeral. The immature state of the technology meant that high-precision imagery was very difficult to recreate.

> I had expectations of [the performer's] hand being a certain way, and when it wasn't I was so frustrated ... Trying to repeat the experience was frustrating because it was never the same again. — Choreographer

The inability to reproduce precision choreography with our reprojected dancers can be attributed to unreliable placement of the virtual scene in the physical environment. The anchoring of the virtual scene into the physical world requires a process known as *localisation*, and we found that the XReal Light headsets used for *Critical Path* performed this process poorly when used within the specific context of a performance environment. Virtual images rendered on the device suffered from a noticeable drift as they failed to properly anchor themselves into the physical space, making it impossible to ensure that virtual imagery appeared in a consistent location between users and over time. This required our choreographer to make some concessions, altering how she would usually approach development of choreography.

> When you wanted to create that type of precision, you also had to have a looseness to know that it will never be exactly the same. It was more about giving the dancer a range. So "this" is the area you can operate in, and if you operate within this, something will happen ... Often [the image] goes through [the mirror]. I loved that. I wanted to do more of that, but it's so hard to recreate ... so in a way if [it happened] during a showing, it was accidental. I mean, I had it in my mind this was a possibility, but I had to stay frustrated and let go of the wish for [this movement] to be precise. — Choreographer

5.4.6. Deployment considerations

While other forms of mixed reality presentation do not match the physicality, dimensionality, and intimacy that AR headsets offer, the price of current-generation headsets must be considered as well as the associated operational expenses, including time and effort required for deployment.

These factors generally limit the presentation of new works to small audiences. Mixed reality performances utilising projection have no trouble delivering an experience to an audience of hundreds or thousands, but purchasing, maintaining and facilitating the amount of headsets required for audiences of scale is simply not feasible with current-generation offerings.

> I think the big limitation is from a theatre perspective ... I am just talking about ... the challenge of multiple headsets ... As a theatre practitioner, you kind of want *everyone* to experience it or see it. I'm immediately up against the limitation of larger audiences ... [It] preoccupies me ... How do we solve that one? — Director

The XReal Light units utilised for *Critical Path*, combined with compatible Android smartphones, cost us roughly $1500 per system. This was an affordable alternative to higher-priced headsets like the Microsoft HoloLens 2 and Magic Leap devices which are closer to $5000AUD per system, however for a lot of teams, it is still prohibitively expensive — especially considering the immature nature of current-generations of this technology.

Headset-based AR requires constant processing, which limits performance and battery life of the devices. Merely possessing 10 headsets does not necessarily allow for 10 participants to use them simultaneously, as the high processing demands of the current-generation technology impose additional restrictions that must be considered (discussed further in 6).

In practice, AR performance on the mobile chipsets was efficient enough to provide users with a 45-minute session — after which, devices would overheat causing both rendering performance to be throttled as well as general discomfort for participants. Headsets heat up slightly on user's faces after prolonged sessions, but the real issue is with the smartphone devices that heat up tremendously in user's hands or pockets becoming incredibly uncomfortable.

> The phone itself was getting so hot … we put it in the pocket and we clipped it so it wouldn't fall, but it was getting too hot in the pants so we had to take breaks. The headsets themselves were getting hot, too.
> — Performer

The limited usage time of the devices meant that we needed to swap between several devices when developing *Critical Path*. We had multiple smartphones in standby, on charge, and would switch these out as devices started overheating or ran out of battery. We got reasonably quick at this, but it did still slow down development. Swapping out a device caused interruptions in the creative process, and triggered a calibration procedure where participants needed to return to a known position of the stage and orient themselves before launching the AR software.

6. System Design

6.1. Introduction

The previous two chapters presented a portfolio of creative works, and discussed the technological system designs that served as the backbones of each. The design and development of each system did not occur at the start of each production, but evolved naturally through the construction of the work in concert with the creative development, building on knowledge collected in each development prior.

The final work of *Critical Path*, as discussed in section 5.4, incorporates a system design that represents the accumulation of this knowledge in the areas of depth cameras, point-cloud manipulation, mixed reality presentation, networking and software design. The software system developed for this work, along with the hardware selection, is therefore split from the creative portfolio and presented in this chapter as a more comprehensive discussion on system design for mixed reality performance. Through introduction of the system developed for *Critical Path* — entitled *pointcaster* — this chapter concludes the technological journey of experimentation and iteration through the entire portfolio.

6.2. pointcaster

6.2.1. Outline

pointcaster enables 3D telepresence by capturing subjects using depth cameras and reprojecting them into mixed reality. It was created to facilitate the ideas explored in *Critical Path*, and as a tool to extend telepresent performances like *Layer* into 3D space. Source code for pointcaster is available at https://git.matth.cc/pointcaster.

The system is comprised of two software components, *pointcaster* and *pointreceiver*.

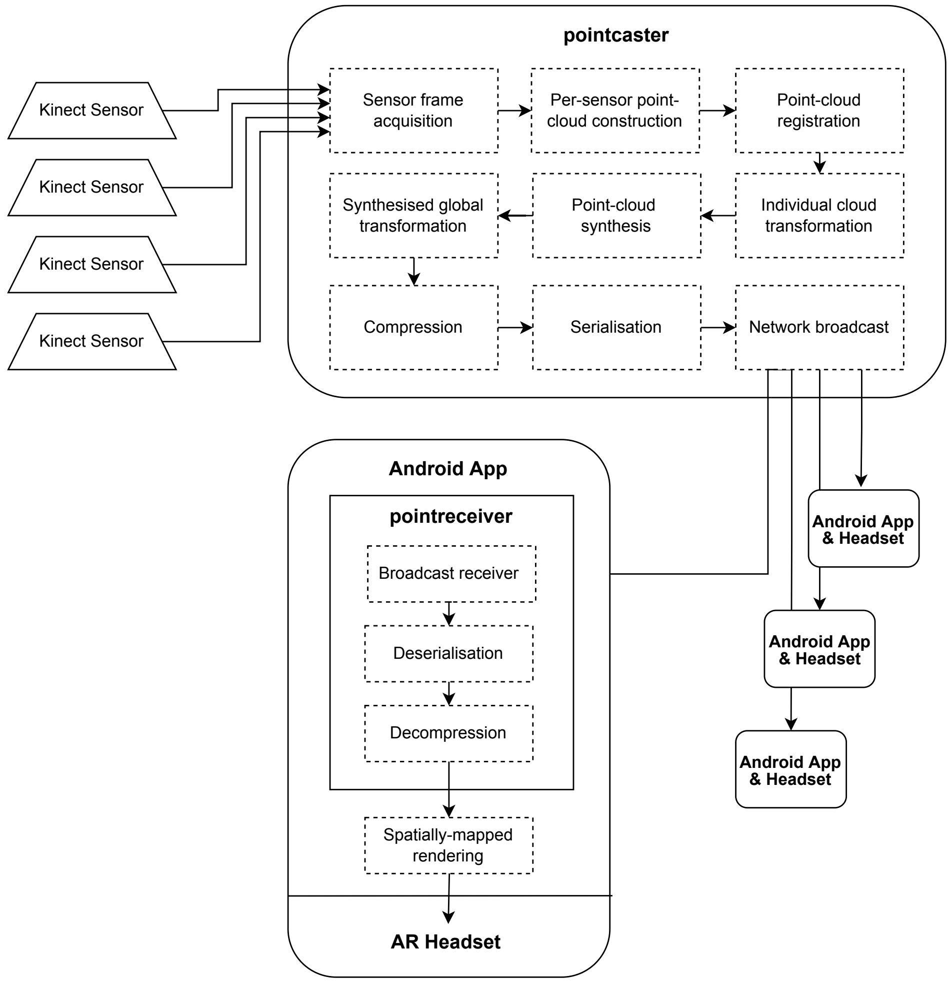
The *pointcaster* server application handles:

- acquisition of point-clouds through connected depth cameras
- aligning the point-clouds of multiple cameras into one (a process known as *registration*)
- transforming the geometry, behaviour and aesthetics of the point-cloud
- encoding the result into a network-transmissible format
- broadcasting the point-cloud for presentation in 3D clients on remote devices

The *pointreceiver* library, embedded into the 3D client applications on remote devices, takes care of:

- receiving the live stream from pointcaster
- decoding the stream to render for mixed reality

Figure 6.1. – *System diagram for* pointcaster.



For *Critical Path*, pointcaster runs on a Linux PC with an AMD Ryzen 9 5950x 16-core processor and an Nvidia RTX 3090 GPU. pointcaster processes data from three Azure Kinect sensors, and broadcasts a point-cloud stream to six instances of the pointreceiver library embedded in a Unity application for Android smartphones. The phones are LG V50 ThinQ devices and are used to drive six sets of XReal Light AR glasses. An overview of the entire system architecture is shown in figure 6.1.

Over previously used telepresence systems, pointcaster:

- integrates high-performance filters and transformation controls desired for creative manipulation of point-clouds
- provides an approach to point-cloud registration that is more suitable for live performance environments
- improves upon point-cloud transmission efficiency through optimised codecs and serialisation

6.2.2. 3D acquisition

The initial stage of a telepresence pipeline involves capturing participants using some form of 3D acquisition system. pointcaster makes use of the Microsoft Azure Kinect platform[1] in order to obtain 3D data in a point-cloud format.

Acquisition technologies

Various capture technologies are used in comparable pipelines for 3D acquisition, including stereo cameras, structured-light-based sensors and time-of-flight sensors. There are pros and cons to each approach, which makes certain technologies more suitable for different ranges of applications.

These 3D capture technologies are packaged into products known as RGB-D cameras — or *depth cameras*, as indicated by the D in RGB-D. A range of affordable RGB-D cameras have been made available over the last decade. Stimulated by the release of the first Kinect sensor from Microsoft in 2010, artists, hackers and hobbyists, as well as researchers and engineers in a range of fields such as robotics, human-computer-interaction and mixed reality have all embraced the commodification of RGB-D cameras.

[1] https://azure.microsoft.com/en-au/products/kinect-dk

Stereo cameras are a form of RGB-D camera that can be combined with computer-vision algorithms to perform real-time 3D scene acquisition for telepresence. Stereo cameras, such as the popular StereoLabs ZED product-line[2], make use of two RGB sensors placed physically apart. Images are acquired from each sensor in unison and are passed into a software pipeline that uses the disparity between the left and right images to calculate a depth map of the scene. RGB stereo cameras — along with setups of more than two RGB sensors that similarly calculate depth from image disparity — have been used for acquisition in various systems for 3D telepresence over the years (Gross et al., 2003; Kauff & Schreer, 2002; Kurillo & Bajcsy, 2013; Sheppard et al., 2007; Vasudevan et al., 2010). However, their use has become less popular since the proliferation of cheap and accurate solutions that combine structured-light or time-of-flight technologies with stereo sensor pairs or even forego them completely.

Though the use of passive RGB stereo cameras for 3D scene acquisition has become less common, stereo sensors are regularly combined with other depth sensing technologies (touched on in the following paragraphs) as the stereo pair provides some benefit over the alternatives. Stereo (or multi-camera) systems for 3D acquisition provide a comparatively high resolution compared to alternatives. The StereoLabs ZED 2 for instance, boasts full-HD (1920x1080) frames, whereas top alternatives at the time of writing have far more limited resolutions. In the case of time-of-flight sensors, the Intel RealSense L515[3] is limited to XGA resolution (1024x768) and the Microsoft Azure Kinect is even lower at 512x512[4]. In theory, stereo systems have the potential to deteriorate less over longer-distances than other technologies, which generally work through projecting light into the captured scene. With alternatives, when a subject moves further from the camera, projected light on the subject becomes more dim and less visible to the sensor,

---

2 https://www.stereolabs.com/zed-2

3 https://www.intelrealsense.com/lidar-camera-l515

4 The Azure Kinect can operate depth acquisition at the higher frame resolution of 1024x1024, although only at lower frame rates unacceptable for telepresence.

resulting in noisier depth images as the camera struggles to reconstruct the scene. As a subject moves further away from a stereo camera, the effective capture-resolution is lowered, though noise shouldn't be introduced at the same rate as systems that rely on projected light, as the subject remains visible to the RGB sensors as long as it can be seen clearly.

The biggest downside of stereo camera systems that rely on RGB sensors is their dependence on visible light for depth capture. While not an issue in controlled settings, unpredictable performance environments, as well as the wide range of lighting conditions that may be present throughout a live performance, makes reliance on visible light quality as captured by an RGB sensor a significant concern. Depth acquisition from these solutions is impacted in both low-light conditions and scenes with a high dynamic range of lighting. An RGB stereo-camera is unable to determine the depth of image components that appear overexposed, or even simply in areas of the image that lack variety in texture, as there may be no apparent difference between images acquired from the left and right sensors in order to calculate disparity information. Though some products on the market today such as the ZED 2 fail to address this deficiency, other widely-used depth cameras that make use of stereo sensors combat this weakness by combining the stereo setup with the use of other technologies such as structured-light.

Structured-light cameras work by coupling one or more camera sensors with a light source that projects an observable pattern of (usually) near-infrared light. The camera's sensors, which are either infrared sensors, or are physically filtered to focus on wavelengths emitted by the projector, capture the projected pattern and calculate depth from its distortion as it appears on the scene (Fofi et al., 2004). Their advantage over RGB stereo cameras is that they do not rely on the quality and intensity of visible light capture in order to produce an accurate depth map. The depth output from the infrared sensor/projector combo is fused with RGB sensors to produce a full RGB-D image.

Products like Intel's RealSense line (particularly the SR and D400 series cameras) combine structured light projection with a stereo pair of infrared sensors, and are some of the most widely used solutions for 3D acquisition today. These cameras utilise similar disparity-based algorithms as RGB stereo cameras to calculate depth, but can operate accurately in low-light and even complete darkness due to the on-board projector. They can also operate well through changes in lighting conditions as the projected light remains constant. Structured-light cameras with a dependency on the infrared or near-infrared spectrum may still be a concern for some performance environments. Halogen lights — still commonly used in traditional theatre venues for stage lighting — emit light in the infrared spectrum, which can interfere with the visibility of structured-light patterns if the projector's wavelength overlaps with the light emitted by the stage light. Similar issues may be caused by use outdoors as the sun's rays could interfere with the visibility of structured-light patterns (Gupta et al., 2013). In practice, the amount of degradation that occurs due to interference from external light sources differs from camera-to-camera depending on factors such as the amount of power available for illumination, the type of pattern emitted, and the wavelengths of light produced by the projector. My own experience with the extent of this problem is discussed further in section 6.2.2.

Along with structured-light based approaches, **time-of-flight cameras** are the other most prominently used technology for 3D scene acquisition. This is the technology used in Microsoft's Azure Kinect and the Intel RealSense LiDAR range (such as the L515). A time-of-flight camera projects light from its own projector onto a scene, but instead of sending out a pattern to observe, the illumination components in time-of-flight cameras project thousands of separate pulses of light. Each pulse corresponds to an individual pixel in the image sensor, and the camera hardware measures the time it takes for each pulse to reflect off a surface in the scene and return back to the sensor. Using this timing information, a depth map is generated on the device. This is in contrast to both stereo and structured-light based approaches where the depth map is generated off the device using software, which places a bigger burden on the computer hosting the device. Time-of-flight cameras therefore save on computing resources compared to other approaches, though in practice, algorithms used for calculating depth from disparity or structured-light are highly optimised, and are generally burdensome only if the host device uses a low-powered (e.g. a mobile) processor or if there is a large number of simultaneous cameras connected to the host. The popular time-of-flight cameras on the market at the time of writing however (including the RealSense L515 and the Azure Kinect), generally have a lower acquisition rate compared to the alternative technologies. While other solutions can reach frame-rates of 60 or 90 frames-per-second, it is rare to see (affordable) time-of-flight cameras that operate over 30 frames-per-second.

Like structured-light, time-of-flight sensors can perform well in low-light and complete darkness, but are generally susceptible to the same external light-based deficiencies as structured-light. The pulses of light they produce are similarly emitted near the infrared spectrum, making them susceptible to other near-infrared light sources. In addition, due to their reflection-based approach, some materials are not captured by time-of-flight cameras. If a pulse from the time-of-flight projector hits an object with a material that scatters light in unpredictable ways, the pulse can be lost and no depth information is attained, or it can be bounced multiple times before returning to the sensor resulting in depth inaccuracies. This kind of inaccuracy can occur when time-of-flight cameras are focused on highly-reflective surfaces, as well as with ”fuzzy” and porous materials (Tölgyessy et al., 2021). For some, this inconsistency through different materials may be an fascinating parameter to explore when devising costume or set-design of live performances that utilise time-of-flight cameras.

I also must mention that the pace of advancement in the field of **machine learning** has been enormous in recent years, and learning-based solutions to a wide variety of computational problems have emerged. 3D scene acquisition has not remained untouched, and novel methods for depth estimation using machine learning models have begun to appear. Some models focus on reconstruction of a 3D scene using the RGB output from a single camera (Godard et al., 2019), and some focus on reconstruction using stereo camera images (Laga et al., 2020). Learning-based solutions to 3D acquisition are exciting because they remove the need for specialised cameras. With the single camera based approaches, artists and researchers should be able to produce RGB-D imagery with whatever camera they already have on hand. While promising, learning-based approaches currently have a few major drawbacks which make them inappropriate for 3D scene acquisition in live performance scenarios. The first drawback is the same issue as with RGB stereo sensors — their reliance on visible light. Secondly, machine learning models require relatively immense GPU resources. So while artists may not need to purchase specialised camera hardware, they do require a high-end (read expensive) graphics card. The computational weight of learning-based solutions also means that models need to be "slimmed down" in order to run at real-time frame-rates, which in-turn makes them less accurate than the non-learning-based solutions already discussed. There is still no learning-based solution that can predict depth with high accuracy using the low computational resources required to run at real-time frame-rates (Masoumian et al., 2022), although the space is rapidly improving and I do not doubt this will be achievable soon.

The various 3D acquisition technologies each offer their own set of advantages and disadvantages. Summarised briefly, RGB stereo cameras provide high resolution, though are highly dependent on visible light and may not perform well in low-light conditions. Structured-light and time-of-flight cameras are less reliant on visible light, but can be susceptible to infrared interference including sunlight and halogen stage lighting, and may not accurately capture certain materials. Machine learning-based solutions although promising, currently require high computational resources and are generally also dependent on visible light. Table 6.1 provides a comparison between these technologies which can help guide the selection of an appropriate acquisition method.

Table 6.1. – *Comparison between 3D acquisition technologies.*

	RGB stereo	Structured-light	Time-of-flight	Machine-learning
Resolution	High	High	Medium	Low
Computing requirements	Medium	Medium	Low	High
Dependent on visible light	Yes	No	No	Mostly yes
Susceptable to IR interference	No	Yes	Yes	No
Low-light performance	Poor	Good	Good	Poor
Material incompatibilities	No	Yes	Yes	No
Conventional frame-rates	High	High	Medium	Low

Practical considerations for 3D camera selection

pointcaster's system for 3D acquisition is based on Azure Kinect cameras. It can host any number of Kinects placed arbitrarily around the performance area. Azure Kinect cameras were chosen for this system due to their features and capture quality. Over the past several years I have explored a range of 3D cameras that helped to inform this decision.

A fair amount of my work has made use of older Kinect v2 cameras. The Kinect v2 and the Azure Kinect both make use of time-of-flight technology, though Tölgyessy et al. (2021) concluded the Azure Kinect produces depth frames with lower noise and higher accuracy than the Kinect v2. The Azure also has a larger usable range of distance subjects can be from the camera. Depth maps the Azure produces contain far less erroneous points than the v2. Depending on the scenario however, these improvements might not be as much of an important consideration for many live performance artists as it would be for those requiring 3D acquisition for scientific or commercial applications. The accuracy of the depth data and level of noise acquired from a Kinect v2 was perfectly adequate for use in a few of my recent performances, and while captured depth resolution is lower than the Azure Kinect, this difference is marginal[5]. I made use of the Kinect v2 in *Trigger Happy Visualised* (section 4.2),and in *Computer Storm* (4.4), *AirStream* (5.2), *Layer* (5.3),

[5] At real-time framerates, the Azure Kinect has a maximum depth resolution of 512x512 — only slightly better than the 512x424 of the Kinect v2.

and *My House Your House* (5.3.6). The Azure Kinect is substantially more expensive than the Kinect v2. When the Kinect v2 was packaged and released for Windows in 2014, it cost $200 USD, and it can regularly be found for $50 USD on the used market today. The Azure Kinect is $400 USD — a sizeable premium that might not be justified if state-of-the-art accuracy and noise reduction is not required.

The Azure Kinect is certainly best-in-class for 3D acquisition accuracy at this price-point, though the biggest practical improvement for choosing the Azure Kinect over the v2 is the increased flexibility for deploying multi-camera arrays. The Kinect v2 is limited to one-camera-per-host-device. This means synthesising multiple cameras into a dense 3D reconstruction involves dedicating a PC to each individual camera in the array. 3D acquisition is made far more complicated to deploy and manage due to the need to create a network to pass frames from each camera back to a server. Systems like *LiveScan3D* (Kowalski et al., 2015) are based around creating a local acquisition network for multiple devices due to this limitation. Additionally, the camera shutters in an array of Kinect v2 devices have no way of synchronising the time at which they capture frames, leading to issues with unaligned point-clouds during fast motion. My initial experiments attempting to use a multi-camera 3D acquisition system based on Kinect v2s highlight the importance for frame synchronisation when capturing dancers in motion. Figure 6.2 illustrates the kind of artefacts produced when frame-capture from multiple cameras on the same subject are not synchronised. Both Azure Kinect cameras and Intel RealSense cameras provide ways of synchronising shutters during capture so that this distortion is reduced.

The productions of *Computer Storm* (section 4.4) and *Layer* (section 5.3) experimented with two different RGB-D camera alternatives from Intel RealSense: the D435, a structured-light-based sensor with a stereo infrared camera, and the L515, a time-of-flight camera.

Figure 6.2. – *The effect of unsynchronised shutters in a multi-camera array. Notice the artefacts around the arms and legs as the dancer moves at speed.*

Due to its stereo infrared cameras, the D435 is able to capture depth at high rates up to 90 frames-per-second[6, 7], though the limited depth-sensing range of the camera severely restricts its usability in live performance. If the D435 is placed any greater than 1.5 metres away from the subject it is attempting to capture, depth images become unusably noisy[8]. Figure 6.3 shows the artefacts produced by the depth sensor in the D435. Notice the camera's inability to determine the shape of the participant's arms and feet at the pictured distance of 2 metres. Even at distances closer than 2 metres, the D435 still produces a noticeable "waviness" that changes frame-to-frame. All three of the time-of-flight cameras I have personal experience with (the L515 and the two models of Kinect) do not exhibit this distortion, and in general are far more temporally stable. Adding to the

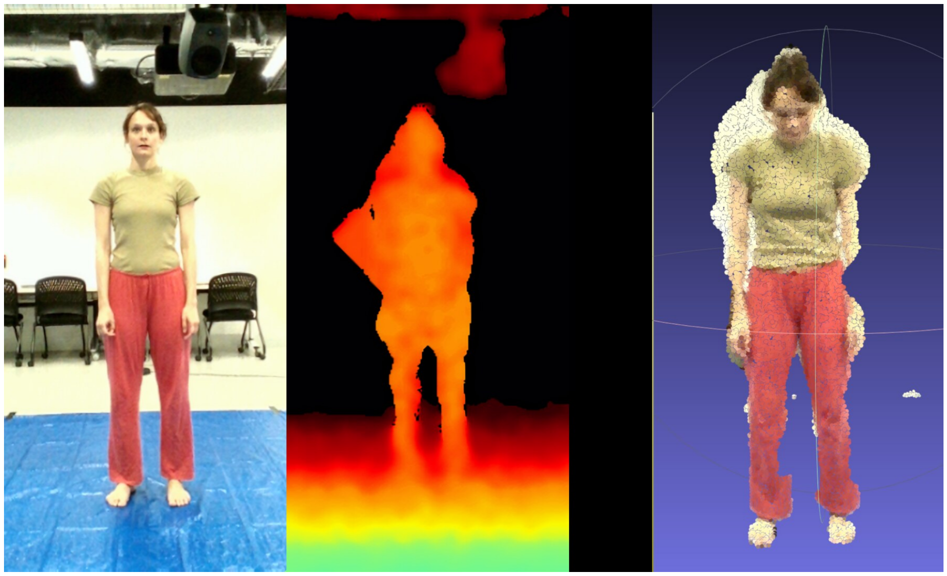
---

[6] https://www.intelrealsense.com/depth-camera-d435

[7] The D435 is actually able to capture depth at rates up to 300 frames-per-second, though only at very low resolutions (256x144). 90 frames-per-second is the highest frame-rate the device is capable of at resolutions useful for telepresence applications.

[8] Intel's marketing for the D435 around its depth-sensing capabilities changed over the course of their use in my research works. The Internet Archive shows in 2019 the product page claimed this camera was capable of sensing depth up to 10 metres. In 2021 this was updated to under 3 metres.

Figure 6.3. – *The Intel RealSense D435 cannot produce accurate depth at a distance of 2 metres. Raw RGB and depth frames are pictured on the left and centre. A point-cloud from a seperate frame is pictured on the right.*



laundry-list of downsides to this camera, it is much more susceptible to infrared interference. If used with halogen stage lighting, it is beneficial to filter the lamp's infrared spectrum with a physical gel.

In comparison, the L515 time-of-flight camera is much less noisy than the D435. It produces good quality depth images at longer ranges, and was used in *Layer* at a distance of up to 5 metres from the performers. Intel have rated this camera for accurate depth at up to 9 metres. Though at that distance the captured point-clouds can be very sparse and reasonably noisy, and some brighter lighting can compromise the depth range even more.

The L515 suffers from the same drawback as most time-of-flight cameras in that it struggles to capture some specific materials. Synthetic materials like polyester behave unpredictably with this camera and are generally missing from the depth images it produces. The L515 also struggles with capturing hair in a way that alternative time-of-flight cameras do not. *Layer* exploited this limitation to artistic effect, but it may not be desirable for some practitioners due to its material incompatibilities.

In addition to hardware considerations, it is also important to evaluate the software that ships with each camera. To incorporate the Intel RealSense cameras into custom software, developers must make use of `librealsense`[9] — a C++ library (and SDK) which provides interfaces for frame capture, point-cloud generation and camera configuration. This library is cross-platform and easy to use, though it distinctly lacks a solution for *body-tracking*. The libraries provided for both the Azure Kinect and the StereoLabs ZED cameras provide simple functions for detecting bodies and retrieving the skeleton of each tracked body. The pose of each joint as deduced from a participant's skeleton is useful in a number of interactive situations where a developer might want a simple way to provide participants natural control over aspects of the simulation or visualisation. Body tracking functionality can also be used to remove unwanted points from a point-cloud that are determined to not belong to any person. In pointcaster, the body-tracking component is used as a simple method to determine camera transforms during *registration* (discussed more in section 6.2.4). Body tracking may still be achievable with the RealSense devices using third-party software libraries, though this is not necessarily trivial.

The Azure Kinect was chosen as the most appropriate camera to use for pointcaster due to its numerous benefits over alternative options. It offers state-of-the-art accuracy and noise reduction at its price-point, with a good depth resolution and acquisition frame-rate. It is performant through a wide range of lighting scenarios, including under halogen stage lighting. Among the depth cameras I have worked with, it has the largest depth range for capturing subjects effectively,

---

9 https://github.com/IntelRealSense/librealsense

and the SDK provides useful features such as body-tracking. While the benefits of this camera places it as an optimal choice, it does still have weaknesses that users should consider when deploying in a performance setting.

Again, like all time-of-flight sensors, capture quality is sometimes affected by certain materials. In the case of the Azure Kinect, it has no issue with hair, but certain clothing is rendered invisible by its captures. Material reflectivity as determined by a material's colour is known to be able to influence the accuracy of time-of-flight technology (Boehler et al., 2003), and we see this in the Azure Kinect, though this phenomenon is not predictable. Figure 6.4 shows my black cotton shirt disappearing entirely in the capture by the camera, though other black shirts and other cotton shirts are captured just fine. Bizarrely, I own another shirt of the same style by the same brand, which looks and feels identical, yet is perfectly captured by the camera. Figure 6.5 shows the two seemingly equivalent t-shirts being captured with different precision. I can't begin to explain this, and publications evaluating how certain materials and colours affect the quality of time-of-flight or Kinect accuracy (Boehler et al., 2003; Tölgyessy et al., 2021) do not explain the inconsistency either. The unpredictability in materials with this camera definitely warrants further investigation.

While inconsistent capture behaviour might not be ideal for some applications, it does open up avenues for creativity when exploring costuming in performance. The ability to choose what is captured by layering different materials could be a fascinating course of inquiry when developing a performance. If a practitioner finds their material is not working how they desire, it is simple to just use an article of clothing that is fully compatible. This is an example of a limitation inherent in the device that can be ignored if desired, but there are some flaws within Azure Kinect that cannot be ignored. One downside in particular that might frustrate practitioners is the Azure Kinect's inconsistent compatibility with different USB cables and ports.

Figure 6.4. – *My shirt as visible on the left does not appear in the capture from the Azure Kinect that appears in the mirror on the right.*

While the cabling is more convenient than the Kinect v2, which requires multiple cables and adapters, managing cables for the Azure Kinect can be fickle. The supplied cable is generally too short for performance scenarios where computers must be off-stage — necessitating the use of long USB cables or USB extenders, though compatibility with these is not guaranteed. Even when using high quality (expensive) USB-C cables that are designed for the highest specifications, some PC motherboards have issues transmitting the required power or data to maintain steady frame-rates. I have tried several combinations of computers, cables, USB ports and USB cards with the Azure Kinect, and reliable compatibility beyond the supplied 1.5m cable is not assured in many system configurations. In cases where the camera cannot maintain a connection, unplugging, re-plugging and re-initialising drivers may be necessary. Generally, once a working combination of motherboard, USB ports and cables is found, and once the driver is successfully initialised, the cameras are stable. I have not had the devices drop out during

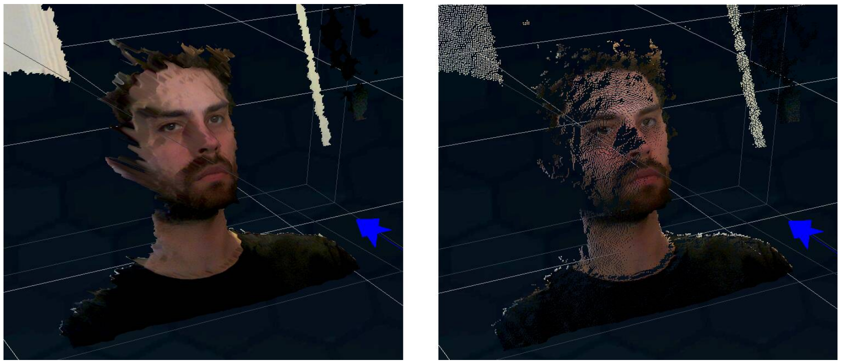
Figure 6.5. – *The difference in quality of capture between the two seemingly equivalent materials is visible even at a close range of 1 metre.*

a running session, but setting up a system based on Azure Kinect cameras can be finicky, and is something that requires forward-planning and lots of testing to make sure PCs and cables are compatible.

Volumetric capture formats

Depth cameras generally provide RGB-D data to users as separate 2D colour and depth frames. To process and visualise the data, it is preferable to convert the separate frames into a unified representation in 3D space. There are two primary volumetric formats used to represent acquired 3D camera data: *point-clouds* and *meshes*. A representation of the same scene as both a mesh and a point-cloud is shown in figure 6.6. The decision to use point-clouds in pointcaster was based on both technical and aesthetic considerations, which I will elaborate on in the following paragraphs.

Figure 6.6. – *A mesh is pictured on the left, and a point-cloud on the right.*



Most simply, point-clouds are a list of coordinates in 3D space. For each pixel present in the depth map obtained by the camera, a point is constructed that describes its position in 3D space. With the addition of a corresponding RGB frame, each point can be assigned its own colour. Each point in the cloud ends up containing the X, Y and Z coordinates, as well as the R, G and B values of the colour to form a structure.

X	Y	Z
R	G	B

A point-cloud is just a list of these point structures, which is what makes it easy to store, manipulate and simply comprehend.

X	Y	Z	X	Y	Z	X	Y	Z
R	G	B	R	G	B	R	G	B

Filtering out points is a simple process of removing the undesired entries from the list, and translating a point's position in space can be done by simply adding or subtracting to the point's coordinate values.

Meshes, on the other hand, are a collection of polygons (usually triangles) connected to each other in 3D space. These interconnected elements create the appearance of a continuous surface, but come at the cost of increased computational and mental overhead. Meshes contain not only a list of coordinates, but also lists of descriptors that outline how the coordinates are joined together into a polygon. Unlike point-clouds, which are *order-independent*, the order of elements in a mesh is crucial for maintaining its structure. Removing undesired components from the mesh is not as simple as removing a coordinate, since the order of the elements within the lists must be taken into account. The same goes for merging data from multiple cameras together. While *mesh fusion* from multiple depth cameras is a very active topic of research and development, I am not personally convinced that creating the appearance of a continuous surface is worth the trade-off in complexity compared to simply using point-clouds.

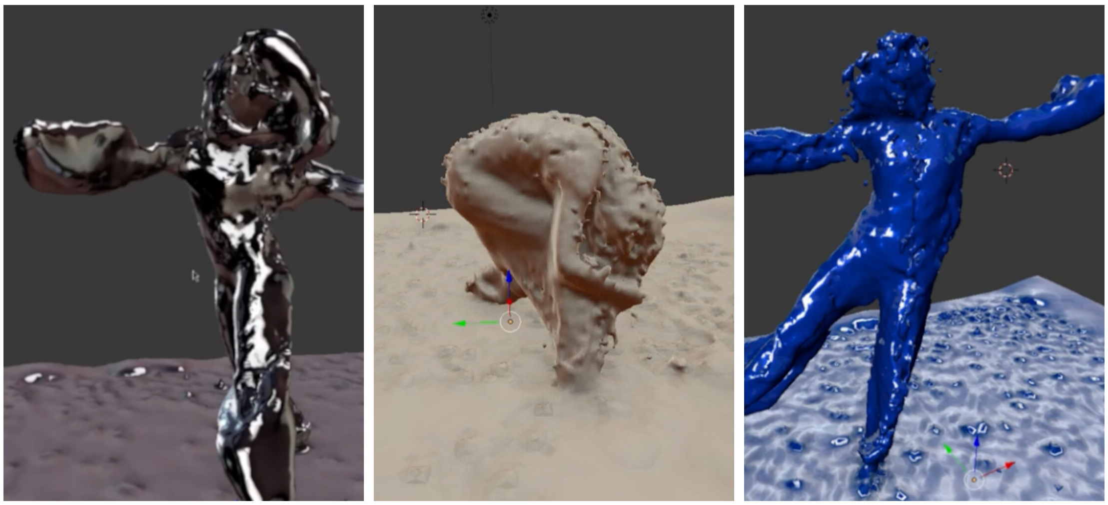
Aesthetically, meshes generated from depth cameras can fall into the realm of the *uncanny valley.* In the context of capturing performers from a distance on a stage, and at a resolution that can be rendered efficiently on mobile AR headsets, coloured meshes created from participants can have the appearance of a 'video game character'. I feel that it can look more "fake" than a point-cloud. Even the highest budget pipelines with dozens (or hundreds) of cameras such as Google's *Relightables* (Guo et al., 2019) or Microsoft's Mixed Reality Capture Studios[10] produce meshes that may appear less natural because of the way that telepresence applications tend to strive towards realism. Our brains are trying to pick the flaws in the imperfect representations that are attempting to be presented to us as reality.

The quality of point-cloud visualisations hint that their source is sampled from elsewhere. The cloud of discrete particles is more comparable to a digital photo, and ironically I feel that the pixelation and the incomplete surfaces that can be present in some point-clouds feels more authentic than the unnaturally smoothed surface of a mesh. My argument here is not that meshes present poorly, but that they can look unsettling if they are presented as too close to real. If practitioners

[10] https://www.microsoft.com/en-us/mixed-reality/capture-studios

embrace the limitations of their technology some beautiful images can be found. Figure 6.7 and 6.8 present meshes created from Kinect v2 devices. There is plenty of creative potential in the mesh format.

Figure 6.7. – *Meshes with their potential for smooth surfaces can wear infinite different materials.*



When using point-clouds in my recent works, I haven't been trying to hide the medium of capture as I feel meshes attempt to. I enjoy the aesthetic quality the point-cloud format brings. It may take some time before volumetric capture technologies produce photorealistic meshes of humans, or it is possible that meshes produced by depth cameras may continue to reside within the uncanny valley. However, dense point-cloud capture produced when standing close to a camera allows artists to create high-resolution imagery that doesn't have video game or uncanny connotations, and sparse point-clouds produced from standing away from the camera or viewing the cloud close up can produce some poetically abstract 'pixelised' imagery. Figure 6.9 shows how viewing the same point-cloud either from a distance or close-up can completely change its quality from photo-like to abstract.

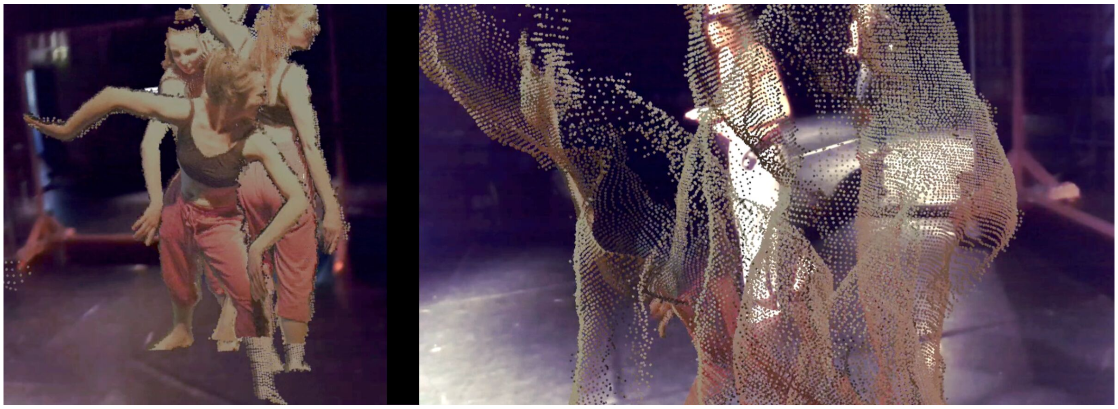
Figure 6.8. – *The performer controlled mesh clones of themself in 'Trigger Happy Visualised'.*

The abstract qualities that can be brought out by point-clouds allowed for a broad range of creative expression and interpretation in the works presented in this thesis. In addition to this, their usage simplifies the technical side of the software pipeline. Both of these reasons are why pointcaster converts incoming RGB-D frames into point-clouds instead of meshes.

6.2.3. Transformation pipeline

After acquisition, pointcaster provides a user interface for manipulating the reconstructed 3D scene once it is acquired from connected cameras. This interface can adjust the position, orientation, scale, colour and density of captured performers,

Figure 6.9. – *Point-clouds captured and layered for a scene in* Critical Path *appear photo-like when viewed at a distance (left), and pointillistic when viewed close-up (right).*



as well as duplicate their bodies and freeze them in time. Computation of these transformations on our volumetric data is handled with a GPU-based transformation pipeline on the pointcaster server. Figures 5.20, 5.24, and 5.25 show examples of the transformation pipeline's capacity for manipulation.

Transforming the large amount of 3D data acquired from the Kinects can be expensive. Transformation algorithms must run through every point in the point-cloud individually and apply the desired manipulations. This can be even more resource-intensive when incorporating many cameras and/or many performers in a development. To allow the transformations to run efficiently enough to operate at real-time rates, this task is handed off to a pipeline of transformation functions accelerated through use of the CUDA GPU-programming SDK[11]. GPU-programming makes it possible to highly parallelise the task of transforming the whole synthesised point-cloud, though it does introduce dependency on a high-end GPU for the pointcaster server.

By placing all responsibility over transforming the 3D scene onto the server, we benefit twofold. Firstly, we allocate less computing load to our lightweight AR client devices. This system requires each participant to wear an AR headset that

[11] https://developer.nvidia.com/cuda-toolkit

performs resource-intensive tasks like SLAM and stereoscopic-rendering (more about this in section 6.2.7). The mobile hardware that powers these AR tasks today is barely efficient enough to compute all this without overheating. This system treats the client devices used by audience-members solely as presentation devices — not much different to broadcasting images to a projector. The Unity application that houses pointreceiver — the library that receives a stream from pointcaster — attempts nothing other than the view-dependent rendering of the incoming stream.

Secondly, the communications from our server to our AR clients is made simple by their *stateless* nature with all point-cloud manipulation occurring on the server. Clients simply listen to the broadcasted stream and attempt to render it.

The goal is to reduce the complexity and processing requirements of the smartphone application driving the AR headsets by placing as much processing as possible onto the server. This is especially important at the time of writing, as the performance of AR using mobile hardware is not yet efficient enough to compute complex tasks like point-cloud transformation alongside the resource-intensive SLAM and stereoscopic rendering they must compute.

6.2.4. Point-cloud registration

pointcaster, and pretty much all works presented in this thesis after section 4.4 use more than one depth camera. Using multiple cameras allows 3D acquisition systems to:

- Expand scene acquisition past the extent of a single camera's frustrum
- Capture multiple perspectives of a subject to create a more comprehensive 3D image

In order to achieve either of these tasks, the incoming data from multiple cameras must be stitched together into a single 3D scene. *Registration* is the process of finding the transformation required between multiple point-clouds for them to be merged into one. The matrix transformation required is the inverse of the camera's physical position and orientation in space. Some options for estimating this transformation include image markers, object detection and body tracking. The state-of-the-art in registration provides many methods of estimation that do not require physical aids like markers or objects (Huang et al., 2021). These methods such as the well-known Iterative Closest Point (ICP) algorithm, or newer machine learning-based approaches are suitable for *fine registration*, after point-clouds are already roughly aligned, where methods utilising physical markers are typically suitable for a *coarse registration*. Existing 3D acquisition implementations *LiveScan3D* (Kowalski et al., 2015) and *Brekel* implement methods for both coarse and fine registration. pointcaster itself implements a coarse registration function using body tracking provided by the Azure Kinect SDK.

With my performances with the Electronic Orchestra Charlottenburg in 2019, I made use of *LiveScan3D* in a multi-camera configuration. Although these performances were a success, the image-marker-based approach to registration used by *LiveScan3D* showcased itself as a weak-point when using it to prepare in the live performance context. For image-marker-based registration, the software takes the image from each connected camera and tries to locate in it a printed *Aruco* marker[12] physically placed in the scene. The software then uses the deformation of the marker in the camera's image to determine the physical location of the camera in space. Marker recognition like this is a reasonably trivial problem in the domain of computer vision, and in *LiveScan3D* it is handled by the *OpenCV* software library[13].

12 Aruco markers are a standard fiducial marker used for spatial tracking in computer vision, first proposed in Garrido-Jurado et al. (2014).

13 https://docs.opencv.org/4.x/d5/dae/tutorial_aruco_detection.html

When a system uses printed markers placed in the capture environment for registration, at least one marker must be clearly visible and appear evenly lit from all camera positions. This is not an issue in controlled environments, but in unknown performance settings (e.g. an unfamiliar venue booked for a performance with a same-day bump-in), it can be a significant issue. One performance with the EOC saw us unable to calibrate three cameras in *LiveScan3D* using the printed markers for registration. Due to the properties of the venue, the system was unable to detect the placed image marker from all cameras at once. A combination of the lights in the venue, reflections on the marker as picked up by the cameras, and the distance the cameras needed to be away from the marker itself in this particular setting, all contributed to the inability to properly register each camera's position before the performance. We had to continue with the cameras unaligned — which was admittedly still a successful performance, though more abstract than initially intended. The image-marker-based approach to coarse registration fails our system requirement of being reliably usable in environments that offer little control over lighting and stage conditions.

Because of this experience with image-marker-based registration, an alternative method was chosen. pointcaster makes use of *body tracking* in the Azure Kinect SDK as a method of point-cloud registration that provides robustness through a wider variety of lighting and unknown stage conditions. The Azure Kinect SDK is able to determine the position and orientation in world-space of people present in the camera's frame. It is able to provide details on the orientation of many joints in the body's skeleton, and it does this using a machine learning model that takes the camera's depth frame as its input. When running registration in pointcaster, the position and orientation of a single user's hips is acquired from each connected camera. Using this known position, each point-cloud is transformed so that the single user's skeleton is aligned — subsequently merging the point-clouds into one.

Unlike image-marker-based registration, which uses colour cameras and depends on the quality of light they receive, this approach based on localisation around a tracked body can operate in a much broader range of lighting conditions, since it uses the camera's time-of-flight sensor instead. This method is also much more flexible in terms of camera placement. Since a printed image in a marker-based approach must be placed flat on some surface to use for registration, cameras must always be placed above that surface in order to have line-of-sight with the marker. With body-tracking-based registration, cameras can be placed in any position and orientation so long as they can see a single user in the space. This is much more forgiving for performance areas with limited space or awkward venue infrastructure.

The enhanced flexibility and adaptability of the body-tracking-based registration as presented in pointcaster sets it apart from existing alternatives. This system allows for the extension of telepresent performances into unconventional or challenging spaces uninhibited by most physical venue layouts and lighting conditions, making it a more suitable solution for use in highly-variable performance environments.

6.2.5. Point-cloud storage and transmission

Once 3D representations of performers are acquired as point-clouds and transformed appropriately, this data needs to be moved onto participant's AR headset devices to be rendered into the performance area. This involves the transmission of large amounts of data, and in order to be used in a live performance this transmission needs to happen fast. To achieve this speed, point-clouds need to be:

1. compressed to reduce necessary bandwidth, and
2. transmitted to AR clients using efficient networking

Raw data taken from Azure Kinect cameras is not optimised for transmission across a network. When data is captured from the cameras and converted for use as a point-cloud (it is not natively captured as a point-cloud, though the Azure Kinect SDK provides functions for this conversion), depth data is represented with 2 bytes-per-pixel, and colour is represented with 4 bytes-per-pixel. At a resolution of 512x512 pixels with a frame-rate of 30 frames-per-second, a point-cloud stream from these cameras has a theoretical maximum bandwidth of 45 megabytes-per-second, or 360 megabits-per-second . Factoring for acquisition from multiple cameras, as well as streaming to many AR participants, the level of bandwidth this data requires is capable of fully saturating a wireless (or even a wired) network. The amount of data sent through a network also directly affects the speed at which it can be sent; more data means higher latency. So to optimise for the multiple users and low latency desired, the point-cloud data must be efficiently compressed before transmission by pointcaster, and decompressed when it is received by pointreceiver.

The body of work relating to teleimmersive streaming of 3D media — also known as *volumetric* media — has grown steadily over the past two decades as XR devices and RGB-D sensors have become more prevalent. As a result of this increased interest, researchers have begun to tackle the problem-area of efficiently transmitting volumetric media. Many solutions have been proposed to encode 3D data into efficient formats for transmission such as Draco[14] and Corto[15], which handle compression of both meshes and point-clouds, as well as LEPCC[16], PCL[17], G-PCC and V-PCC[18], which focus specifically on point-cloud compression.

[14] https://google.github.io/draco

[15] https://vcg.isti.cnr.it/corto

[16] Limited Error Point Cloud Compression (LEPCC) — https://github.com/Esri/lepcc

[17] Point Cloud Library (PCL) — https://pointclouds.org

[18] Geometry-based Point Cloud Compression (G-PCC) and Video-based Point Cloud Compression (V-PCC)
— Graziosi et al. (2020)

G-PCC and V-PCC are both standardised by the Moving Picture Expert Group (MPEG), and are considered best-in-class when it comes to both visual quality *and compression ratio — that is, the amount that the size of a point-cloud can be reduced by encoding. Publications benchmarking various codecs show that G-PCC and/or V-PCC outperform the alternatives in terms of these factors with /dense* (i.e. high-resolution) point-clouds (Bui et al., 2021; Wu et al., 2020). Results are less impressive when point-cloud data is *sparse* (i.e. low-resolution), and the relatively long time these algorithms take to both encode and decode makes them unsuitable for telepresence applications that require real-time framerates (van der Hooft et al., 2020). Bui et al. (2021) found that Draco encodes colour point-clouds anywhere from 8-17 times faster than G-PCC, and 200-400 times faster than V-PCC. The MPEG algorithms offer an order-of-magnitude higher compression ratio, however Han et al. (2020) found that the Draco codec is still able to reduce bandwidth by a factor of 4.5-6 times over the source data, and was the best-performing library out of their comparison with two other open-source point-cloud codecs fast enough to be used for real-time applications (PCL and LEPCC).

The *LiveScan3D* system published by Kowalski et al. (2015) — which I have used for a number of the presented works — utilises the Zstandard (zstd) algorithm[19] to compress RGB-D frames from Kinect v2 cameras before streaming over the network. zstd is is a general-purpose compression algorithm, and while encoding and decoding speed is fast, it does not achieve compression ratios comparable to codecs specifically built to target point-clouds. pointcaster initially implemented compression using zstd, though benchmarks performed during development revealed (unsurprisingly) that Draco was more efficient than zstd in terms of compression by at least 3x, and up to 10x for denser point-cloud frames. Newer releases of *LiveScan3D* targeted for Azure Kinect cameras utilise Zdepth compression[20] — an algorithm that specifically targets the RGB-D frames coming from the Azure

---

[19] https://zstd.net

[20] https://github.com/catid/Zdepth

Kinect by integrating a quantisation function based on the inherent properties of the camera itself. Even though pointcaster makes use of Azure Kinect cameras, the specialisation of this algorithm makes it unsuitable, since pointcaster performs compression *after* the Azure Kinect frames have been converted into a point-cloud, aligned through registration, and arbitrarily transformed through user-controlled manipulations.

Ultimately, Draco was chosen as the codec for pointcaster due to its fast encoding and decoding speeds, satisfactory compression ratios, and suitability for real-time telepresence applications. In addition to this balance between performance and compression efficiency, Draco's easy-to-use C++ library makes it highly appealing to developers and simple to integrate.

After compression, in order to move point-cloud data off of the server and onto the AR devices, pointcaster makes use of ZeroMQ[21] — an efficient socket communications library that offers low latency and is designed to handle a large number of simultaneous clients. It is generally considered to be best-in-class in regards to message throughput and is used by a wide variety of applications that require low-latency (Hintjens, 2013). This is the same messaging system used for network communication by *Sensor Stream Pipe*[22] (SSP) — the system for RGB-D streaming I made use of in *AirStream* (5.2). The flexibility of ZeroMQs messaging patterns made it an ideal choice to use in pointcaster. In the context of real-time telepresence applications, the priority lies in maintaining low latency rather than ensuring delivery of every frame. If the application is unable to send a frame within the desired real-time rate, the frame should be dropped. As such, pointcaster uses ZeroMQ's `RADIO-DISH` messaging pattern, which establishes the pointcaster server as a "radio" that broadcasts to many "dish" pointreceiver clients who simply listen to the point-cloud broadcast and display it. Akin to an analog radio signal, if one of the AR clients is not "tuned in", or for some reason misses a slice of

[21] https://zeromq.org

[22] https://sensor-stream-pipe.moetsi.com

the broadcast, the behaviour of the server doesn't change. At the application level, the server does not care who is listening to the broadcast, and does not attempt to make any effort to ensure messages are delivered. In contrast, the implementation of ZeroMQ in SSP utilises the `PUSH–PULL` messaging pattern, which is more suited to an inverse application where mobile clients offload their own data onto one or many servers for processing. It also guarantees that every frame is transmitted to every client, which is a behaviour that is detrimental for telepresence over unstable networks. If frames are unable to be delivered due to a momentary decline in network conditions, it is preferable to simply discard these frames and prioritise transmission of the latest data so that we continue as close to real-time as possible. By optimising the messaging pattern to prioritise the latest frames and the lowest latencies, pointcaster ensures efficient transmission of point-cloud data more suited for telepresence.

Despite the efficiency of the combination of Draco and ZeroMQ, pointcaster's network transmission system ensures latency is proportional to the amount of data sent from each individual camera. In essence, the closer a performer is to a camera, the longer the delay between movement and visualisation due to the increased point-cloud density that Draco must compress. This is reasonably predictable behaviour, but adds an extra cognitive burden on the performer as they must account for shifts in latency dependent on the spatial arrangement and choreography. Professional performers are generally adept at adapting to consistent latencies in their performance, such as musicians adapting to the natural latencies of concert halls. The inconsistent latencies dependent on the performer's position in front of the camera though might make their job trickier. It is possible that having a consistent latency, even if it means an increase in the minimum latency, might be a more suitable solution for performance scenarios, although further investigation would be needed to confirm this. None of the performers commented negatively on the experience of latency, though it would make for an interesting Quality of Experience study to pursue further.

6.2.6. pointreceiver and portability

Once point-clouds are acquired, registered, transformed and broadcast by pointcaster, they are received by instances of the pointreceiver library. This library creates a connection with pointcaster to receive the point-cloud broadcast, decodes the incoming data, and passes it onto the application which is responsible for presentation.

It is important that the reception of the broadcast and decoding of the point-cloud is handled separately from the software responsible for rendering and presentation in order to maintain pointreceiver's utility into the future.

The industry has not yet settled around a standardised solution for augmented reality presentation compatible across different manufacturers and platforms. As such, it would be a mistake to bind the functionality of pointreceiver into the same software responsible for rendering the point-clouds into the environment.

The augmented reality headsets used for *Critical Path* are a product called XReal Light[23]. This is a first-generation product manufactured by an emerging company. Many of the companies producing commercially-available augmented reality headsets are new startups. In the case of larger manufacturers, devices are often marketed as development kits for initial exploration and research purposes.

Given the immaturity of the augmented reality headset market, who can say whether these specific solutions and the services they rely on will remain relevant or even operational into the future? The XReal Light headsets run applications

[23] https://www.xreal.com/light

built using Unity[24]. This offers *some* degree of portability across hardware platforms, but scenes built in Unity are obviously tied to the proprietary Unity software ecosystem, and use with these headsets require an SDK and library tied specifically to XReal.

To port an augmented reality experience to work on a different device, the application must be re-implemented with the new device-specific SDK. By investing heavily into a specific vendor's solution, we risk becoming locked into a technology that could soon be superseded, deprecated or even abandoned. The rapid rate of technological change over the course of my PhD candidature has demonstrated this risk.

Some technologies that have the potential to mitigate this issue are *OpenXR*[25], which provides a cross-platform application interface for supported devices, and *WebXR*[26], which goes a step further and aims to offer delivery of applications through WebXR-enabled web browsers available on all mixed reality devices.

These technologies will help future-proof and simplify the development process of new mixed reality experiences by offering simple deployment to a range of devices, cross-vendor integration with minimal code changes, and wide compatibility across smartphones, desktops, VR and AR headsets. The development of these technologies is widely-backed by industry, and while momentum is gaining around these open-standard solutions, both technologies are still relatively immature and proprietary SDKs such as ARKit[27] and SnapAR[28], along with proprietary engines like Unity and Unreal see more widespread usage as of today.

---

[24] https://unity.com

[25] https://www.khronos.org/openxr

[26] https://immersiveweb.dev

[27] https://developer.apple.com/augmented-reality/arkit

[28] https://ar.snap.com

Given these considerations, pointreceiver and its implementation in *Critical Path* on XReal Light devices uses as little platform-specific code as possible. All functionalities unrelated to the concept of presentation and rendering are bundled into a reusable component library with an API written in C to allow for maximum portability between platforms.

Aside from implementation of pointreceiver into the Unity-based applications for XReal headsets in *Critical Path*, the library has also been 'wrapped' for use in TouchDesigner. Similar lightweight wrappers for any desired software rendering system can be created with relatively minimal effort, reinforcing the strategy of minimal platform-specific code for maximum portability and longevity.

6.2.7. AR Headsets

Since pointreceiver is portable between platforms, it is usable inside projected mixed reality productions and other forms of 2D or 3D displays. Its first and primary use case however, is to enable 3D telepresence through display in augmented reality headsets. This section discusses the use of augmented reality headsets in the context of live mixed reality performance and explores the requirements and considerations introduced through their integration.

Synchronisation and network requirements

Presenting mixed reality using projectors usually requires the virtual scene to be rendered only once, and distributed to a small number of video outputs. In the case of *Computer Storm* (4.4) the scene is rendered and sent to a single projector. For

Layer (5.3), the teleimmersive environments were rendered into *two* unique visualisations. The use of AR headsets however, bring about an increased complexity to this process.

The graphics in *Critical Path* (5.4) for example, is rendered in six headsets unique to each audience member, necessitating the orchestration of six instances of the visualisation. To provide audience members with a shared experience, the virtual scene must be delivered to and synchronised between each instance on all headsets.

Multiple users sharing the state of a single virtual scene is a common requirement in systems for multiplayer gaming, which means multiplayer game engines have partially solved the problem of virtual scene synchronisation for us. However, the desire to share 3D camera feeds for telepresence is not commonplace in applications like gaming, so is not fully realised by existing systems. Compared to presenting mixed reality performances with a projector, the complexity in the software and networking infrastructure required when using AR headsets is notably increased.

The transmission of point-cloud data to a network of mobile AR headsets can be achieved via Wi-Fi, but the use of a wireless link places additional constraints on the network compared to streaming between a wired network of computers. A high network throughput is desired to deliver an adequate resolution point-cloud, and as previously mentioned, a low-latency is desired for increased presence of participants.

Despite the latest Wi-Fi 6e standard being theoretically capable of operating at speeds of 9.6 gigabits-per-second (Oughton et al., 2021), the best case scenarios see speeds less than a tenth of this in practice (Florwick, 2022, July 14). Practical constraints such as distance of users from access points and the number of connected devices can dramatically reduce Wi-Fi performance.

The dependence on Wi-Fi in order to provide users unrestricted movement (free of cables) reinforces the need for efficient compression and transmission of point-cloud data as elaborated on in section 6.2.5. Depending on the scale of the performance area, the number of participants, and the number of depth cameras used, practitioners may need to consider deploying a non-trivial wireless network consisting of multiple Wi-Fi access points in order to deliver acceptable latencies and throughput to each headset.

Managing physical deployment

In addition to the added network requirements of headsets, management of the physical devices themselves must be considered. Augmented reality processes such as SLAM require a relatively large amount of computation for a mobile device, and these tasks are not yet optimised at the hardware level like decoding a video might be. This high computational requirement results in rapid battery drainage and can generate a significant amount of heat. Practitioners deploying headsets in a performance environment must manage these two factors accordingly.

Over time, the chips used to drive the AR headsets respond to the generation of heat with *thermal throttling*. This safety feature reduces the speed of a CPU when it gets too warm, effectively lowering its performance to allow the device to cool down. This uncontrollable decline in the chip's performance places a restriction on the duration of a production, as thermal throttling will negatively affect the performance of all aspects of the device.

Wearing AR headsets for a sustained period of time can also result in discomfort for the wearer. These devices have the potential to cause eye-strain, and all of the people involved in developing this work noted at least minor levels of discomfort due to eye-strain after prolonged sessions of wearing the headsets for multiple hours.

Nausea symptoms and motion sickness typically associated with VR headsets — known as *simulator sickness* — is known to be less pronounced in AR (Vovk et al., 2018), and we found that none of the development team and no audience members for *Critical Path* reported symptoms like this. A certain level of discomfort can also come from the heat generated by the headsets.

The extent to which these issues affect each wearer can vary significantly. So while these devices are not likely to cause simulator sickness in the wearer, eyestrain and the effects of heat on comfort do need to be managed.

Compared to scrim projection

Five out of the seven works presented in this thesis utilised projections in order to realise mixed reality spaces for performers and an audience. Out of these, four were projected onto a transparent theatre scrim to create an illusion similar to a hologram, suspended in front of the performers on stage. In this body of work, scrims were used to place virtual audio-visual objects (4.3), augmentations of the stage (4.2), and remote live participants (5.3.6) into the performance area. The scrim is a fantastic medium that allows for a range of mixed reality interactions. However, transitioning to AR headsets ameliorates some of the considerations and drawbacks of scrim projection.

One key issue with scrim projection is the need for well-considered lighting in the performance area. The performance environment must be lit well behind the scrim in order for the performer and the stage to appear clearly behind it. This becomes more of an issue when lighter coloured scrims are used, so the colour of the scrim itself is also an important consideration. White scrims, which allow for bright, clear images on low-powered projectors, are understandably less capable of being rendered 'invisible'. With enough brightness on stage, lighter coloured scrims can appear invisible, but with any amount of ambient light in the audience, these instantly lose much of their translucency. In general, the brightness of stage lighting behind a white scrim needs to be greater than the light coming from in front of it. This makes the system less portable from venue to venue, as some venues (such as clubs and pubs) might require ambient light in the audience area at all times. Even if a convincing invisibility is created with a white scrim by having the audience in darkness and the stage well lit, this configuration makes it extremely challenging for a performer to see through to their audience, making them unsuitable despite their bright imagery. Black scrims are ideal for invisibility, and are the least intrusive on lighting design as they reflect the least amount of light — though their use requires very bright (expensive) projectors. Grey scrims were chosen for use throughout my work, as they are a good compromise between the two extremes, though there were many cases where invisibility of the fabric was unattainable. The use of AR headsets removes much of the detrimental effect stage and environment lighting can have on the virtual components in the scene. They are able to maintain a full connection between the performer and their audience, and do not present an immersion-breaking object at front-of-stage if lighting is not quite right.

The main difference between scrim and AR however, is perspective. When virtual images are projected onto the 2D surface of a scrim in front of the performer, the illusion of co-location between the virtual and physical occurs only if the audience is positioned directly in-front of the surface where virtual elements are projected. Viewing the performance at an angle immediately breaks the illusion if a scrim is used. While this doesn't completely spoil a performance, it changes its nature so the audience is more likely to perceive performers interacting with virtual elements that are clearly two-dimensional. On the contrary, AR headsets use a variety of techniques to dynamically update what's presented to each wearer in an effort to maintain the semblance of three-dimensions.

By providing the audience with AR headsets to wear, we can more effectively craft the illusion that virtual elements in the mixed reality environment share the same physical space with performers. This allows works utilising headsets for mixed reality to be presented successfully in less restrictive performance environments.

Much of the work presented in this thesis was designed for presentation in a traditional proscenium arch stage setting, which had the by-product of limiting differences in perspective available between audience members. The use of AR headsets for *Critical Path* enabled us to craft an intimate experience where the audience is free to move and explore the performance environment without disrupting the illusion of mixed reality.

At the time of writing, AR headsets have only been available to consumers for half a decade, and have been generally marketed as *development-kits* targeted to early adopters. They expand the possibilities of mixed reality performance greatly by offering views into the virtual scene from any perspective, but the technology is expensive and immature. It is likely that issues such as power-efficiency and comfort will improve through the next generations of hardware — we'll likely also see a reduction in cost — but the overall added complexity to deployment in comparison to transparent projection screens shouldn't be ignored. A comparison between these presentation technologies is summarised in table 6.2.

Table 6.2. – *Comparison between MR presentation technologies for live performance.*

Projection on scrims	AR headsets
• Limited to a single viewing perspective	• Works with any viewing perspective
• Limited options for audience positioning	• Audience can be anywhere in relation to performer & visuals
• 2D projection provides limited physical/virtual co-location	• Stereoscopic 3D more convincingly grounds virtual elements in space
• Can lack dimensionality, 'flat' by nature	• Strong dimensionality, sculptural quality
• 'Invisibility' affected by theatrical lighting conditions	• Illusion not affected by light, but SLAM can require even lighting

Continued on next page

Continued from previous page

Projection on scrims	AR headsets
• Simple software requirements	• Added software & network complexity
• No disruption in audience comfort	• Comfort dependent on individual
• Unlimited performance duration	• Hardware limits performance duration
• Suitable for large audiences	• Suitable for small audiences
• More incorporable in conventional performance presentation	• Novel mode of presentation opens up new creative possibilities

Comparison of augmented reality headsets

During my research, I had the opportunity to work with two different augmented reality headsets: the XReal Light and the Microsoft HoloLens 2 (pictured in figure 6.10). Both devices have unique characteristics and perform differently under varying circumstances.

Participant's ability to freely move and express themselves without being hindered by the weight or bulk of a headset is crucial for the development and presentation of live performance. The XReal Light is a smartphone-based system, which must be plugged into a compatible Android device to drive it. One of the key advantages of a system like this in the context of live performance development is its lightweight nature, weighing in at only 106 grams. However, this physical benefit was somewhat compromised due to the requirement of tethering the glasses to a smartphone. With the XReal Light, we often had the smartphone tucked into pockets or held in a bumbag. The cable between the smartphone and headset restricted movement to an extent, and during intense movement the smartphone was prone to falling on the ground. This could definitely be remedied by strapping the smartphone to a performers body, though a custom pouch might be worth consideration as we often required access to the device during development to upload changes to the software or reset it.

Figure 6.10. – *The XReal Light (left) is powered by an Android smartphone, while the Microsoft HoloLens 2 (right) contains a fully integrated computer.*

The HoloLens 2 in comparison, is a fully integrated system that houses all computing components necessary within the device itself at the cost of added weight and bulk. It is *over five times heavier*, weighing in at 566 grams. Despite the mobility restrictions of the XReal Light, the bulk and weight of the HoloLens 2 proved to be much more limiting.

Another point of divergence between the two systems is in their SLAM performance, and subsequently their capacity for object stabilisation. In this aspect, the HoloLens 2 is superior, providing a more immersive mixed reality experience by stably locking virtual objects into the physical space. The XReal Light's capability in this area is noticeably weaker, with objects tending to shift and slide around space more. This is particularly noticeable when the wearer is moving around fast. While the use of XReal Lights in *Critical Path* showed that their stabilisation is passable (no participants even mentioned this as an issue since they generally had no point of comparison), they exhibit much lower SLAM performance than the HoloLens 2.

The software development process for each system is more-or-less the same. Each make use of the Unity game engine for building virtual environments, where a component library is easily imported into a Unity project, though the HoloLens 2 can also support Unreal engine. Both software libraries include similarly performant hand-tracking, which allows for intuitive interaction from participants if it is desired. Though I have not implemented this in any of my own work, as interactivity is provided through depth cameras.

One last important note is the performance of each headset's displays. On both headsets (perhaps due to a limitation of the display technology) gradients and bloom effects tend to blow out, leading to pronounced colour banding. An intuitive creative direction for visualisation in these headsets is to add some glow to the virtual objects — in line with the classical depiction of a hologram. Though without careful adjustment of a glowing or blooming element's gradient falloff, the effect can look quite poor. Some bloom effect was used for the visualisations of point-clouds in *Critical Path*, but it needed to be subtle.

Both of these devices have their strengths and weaknesses, with the biggest differentiator between the two being their physical design. For completeness, a table is presented below summarising the comparison between the two devices.

Table 6.3. – *Comparison between the XReal Light and Microsoft HoloLens 2 AR headsets.*

XReal Light	**Microsoft HoloLens 2**
• Lightweight (106 grams) • Tethered to smartphone • Partial restricted mobility due to tethering • Adequate SLAM performance • Supports Unity engine for development • Adequate hand tracking • Costs around $1500 AUD	• Heavy (566 grams) • Fully integrated • Greater restricted mobility due to bulk • Superior SLAM performance • Supports Unity and Unreal engine for development • Adequate hand tracking • Costs around $5000 AUD

6.3. System design heuristics

Adjacent to the development of the seven new mixed reality performance works presented in this thesis was the development of seven technological systems to enable these performances. I have presented key learnings discovered in the practice-based design and implementation of these systems throughout the creative portfolio, and elaborated on them again in the context of the latest system design in the previous section on pointcaster.

Though my journey of system design continues to unfold, certain key technological considerations have crystallised as important concerns across the range of possible mixed-reality implementations in the context of live performance. These considerations, presented in this section, are organised as a set of four *system design heuristics*, designed to instruct future developments in this field. These heuristics — identified and refined through experiences with my artistic collaborators, the practical realities of system deployment, my own personal vision and the broader global context — encapsulate insights drawn from the evolution of my system designs over the past five years of performances.

The four system design heuristics identified are: *adaptability*, *resilience*, *expandability*, and *creative control*. These criteria, which have emerged as critical concerns in the creation of mixed reality systems for live performance, are discussed below. Each heuristic is first explained, and then examples are presented of their effect on system design in the context of the productions presented in this thesis.

6.3.1. Adaptability

The context of a live performance is never one-and-the-same. In order for a system to be useful across multiple performances it must be *adaptable.* This heuristic is concerned with the breadth of circumstances in which the system can be successfully used. It refers to a system's capacity to be deployed in a wide variety of physical contexts where dance, music and theatre take place. A system designed from the start to work reliably in both traditional and non-traditional performance spaces can eliminate hours or days of development time dedicated to deployment in a specific venue.

Key factors to consider when designing for adaptable mixed reality performance systems are *lighting constraints*, *venue physicalities* and *uncontrollable infrastructure*.

- The system design for *Computer Storm* (4.4) presented challenges due to its inability to adapt to lighting constraints of the performance environment. The cameras used in this work were unable to adapt to the low-light conditions present in the final performance area. As a result, performance of 3D capture dropped by half as the cameras adjusted their shutter-speed to let in more light. An ideal system would allow for fine-grained control over camera behaviour.
- The use of *Aruco* markers for multi-camera registration in works utilising *LiveScan3D* seemed adaptable to physicalities of many venues at first, but proved troublesome when depth cameras in some performance environments refused to identify and track these markers. The physicality of the venue forbid the placement of cameras in locations that could properly sight a centrally-positioned marker. The transition to the use of skeleton-tracking based camera registration in pointcaster makes the system adaptable to a wider range of venue physicalities.
- The performance of the *Virtual Audio-Visual Instruments* on the AirSticks (4.3) at the Guthman competition presented an unforeseen issue related to the venue-supplied projector — a delay in the video signal that negatively affected the temporal integration of virtuality and gesture. This element of the performance

environment was beyond the influence of myself or the performer, and although this kind of uncontrollable infrastructure is hard to predict, the performance would have benefited from a system design compatible with increased video latencies.

6.3.2. Resilience

When constructing systems for performance, varying (sometimes unexpected) conditions of the environment must be considered. Where adaptability relates to deployment across different performance environments, *resilience* refers to a system's technical reliability under changing conditions within a single performance. This heuristic is concerned with a system's ability to handle disturbances and return to a normal functioning state.

A system must be resilient so that it does not fail during use, or otherwise inhibit the creative expression of a performance's participants.

- Lighting in a performance will change colour, brightness and position for aesthetic and dramatic purposes. In the presentation of *Critical Path* (5.4), pulling the lights down low — as the director wanted to naturally do between scenes — caused the SLAM in the AR glasses to fail. The device's understandings of the environment were corrupted and the virtual scene would shift in location until the application was reset. Even if the technology of a headset is dependent on visible light, resilience of its system should be programmed in by means of additional tracking or a robust spatial memory.
- The design of pointcaster (6.2) prioritises resilience. It uses a radio-style broadcast to stream point-clouds, which participant AR headsets can simply 'tune in' to receive. If a headset crashes or runs out of battery during a session, it can be quickly replaced. Once the headset is loaded, it immediately begins receiving

the stream, minimising any disruption to the performance. While this approach doesn't prevent crashes from happening, it does ensure that such incidents have a manageable impact on the overall performance.

- The use of AR headsets tethered to smartphones is not resilient to the regular range and speed of motion a dancer might use when performing. In *Critical Path* (5.4), this restricted the creative expression of the performers so much that most choreography was composed with one dancer performing and the other watching on a headset. A next-generation headset that is as lightweight as the ones used, but unencumbered by a cable connecting to a smartphone could be more resilient to a performer's unaltered practice.

6.3.3. Expandability

Expandability refers to a system's ability to handle a growing amount of work, and its potential to be enlarged to accommodate that growth. A system designer might eventually want to cater to an increasing number of participants, a larger physical performance environment, a longer performance duration, or larger amounts of data to process.

It also refers to the ability for the technologies in a system to be expanded in the future through *upgrades*. The market for mixed reality products like AR headsets, depth cameras, and the software to drive MR experiences is immature. If any longevity is desired in a mixed reality system, it must be developed through the understanding that current technological approaches may not maintain suitability through the lifespan of the system.

- *AirStream* (5.2) streamed point-clouds from two depth cameras over a network. When using this system to deliver point-clouds constructed from more than two cameras into a single virtual environment, latency would increase dramatically.

Pointcaster was designed with expandability in mind. Networking optimisations facilitated the addition of more camera streams, allowing the system to be expanded to more participants and larger environments.

- The multi-camera system used for 3D reconstruction in *Computer Storm* (4.4) made use of depth cameras that required their own host computer per device. To add more of these cameras for wider capture areas and more participants would demand management of an unsustainable increase in computers. Too many would become overwhelmingly complex to deploy and relocate, inhibiting expandability.
- pointreceiver is a small C-language library built to receive point-clouds from pointcaster and display them in a mixed reality environment. Its format makes it embeddable in nearly all systems, which simplifies the expansion onto future devices. As a device made by a startup, the XReal Light glasses used in *Critical Path* (5.4) have a reasonable chance of becoming unsupported in the near future, and the pointreceiver library facilitates an expansion path if they become no longer available.

6.3.4. Creative control

The final heuristic in the set is *creative control*, which describes the ability of a system to enable diverse creative capacities. Mixed reality systems might offer high levels of creative control if their technology facilitates varied modes of interaction and a wide range of aesthetic possibilities.

Creative control concerns the interaction design of a system by considering how the technology adapts to and pushes back on an artist's practice. When approaching this concept, a system designer might ask if the system is both flexible in its programmability, but also consistent enough to enable *repeatable* creative exploration such that an artist can build intuition around its performability.

- The system built for *Trigger Happy Visualised* (4.2) offers minimal creative control. It was constructed to present mixed reality visualisations of a pre-defined collection of music. This system is interactive, but its lack of higher-level manipulation limits its capacity for a broad range of creative expression. In contrast, the *Virtual Audio-visual Instruments* for the AirSticks (4.3) can change aesthetics and behaviours at the whim of the user to construct new performance ideas.
- The SLAM tracking on the headsets used in *Critical Path* (5.4) was not sufficient to allow for the precise placement of virtualities within the performance environment. The choreographer's compositional agency over a virtual performer's location on stage was curbed. The lack of definite control over the system as a result of the unstable technology diminished the user's creative expression.

6.3.5. Conclusion

Designing a performance system for repeated and extended use that can incorporate the diverse artistic practices of multiple collaborators over different performances is a complex endeavour. Successful systems for computer-aided performance are not constructed to test an idea once or to function within the confines of a lab environment. The technical learnings explored in this chapter, and the resulting system design was informed through a series of real-world deployments and tested under the demanding conditions of live performance and exhibition. The heuristics of adaptability, resilience, expandability and creative control have been reduced from this experience, and present criteria that can serve as beneficial guidelines for designers looking to implement robust and useful systems for live performance.

It is worthwhile to acknowledge however, that the task of completely satisfying heuristics like these is almost Sisyphean in nature. Practitioners like myself that employ technology as a creative tool are constantly attempting to push such technology to its limits. With each new performance, and each new iteration of design, the desire to expand one particular aspect of a system's potential often results in a concurrent reduction in another. This tension however, is not a deficiency. It is in the essence of our journey as creative practitioners.

While these heuristics offer guidance in the approach to designing systems for mixed reality performance, they should not be understood as rigid parameters to fulfil. They represent concepts that can be used to help navigate, negotiate, describe and evaluate the complex task of system design, and are presented in this thesis as a contribution to aid in future development of mixed reality systems for live performance.

7. Modes of Interaction

7.1. Introduction

Having just discussed criteria aimed at guiding the design of mixed reality performance systems from a perspective of technical implementation, this chapter instead focuses on the theory of interaction enabled by such systems. Setting aside technological details, four *modes of interaction* are presented as an exploration of how mixed reality performance systems mediate the interaction between physical and virtual components.

Applying a theoretical lens to the examination of design artefacts can guide practitioners in creating novel interactive systems. In the context of theoretical *interaction models*, Beaudouin-Lafon (2004) claims a methodological view into interactive systems provides three capabilities: A *descriptive* capability for defining the interaction of systems, an *evaluative* capability for comparing different design alternatives, and a *generative* capability for promoting the creation of new designs. In a similar vein, the modes of interaction presented in this chapter embody the same capabilities. These four modes offer a structured approach to describing and evaluating how physical and virtual elements can interact in a mixed reality performance system, and provide guidance for the generation of future work.

7.2. Existing models

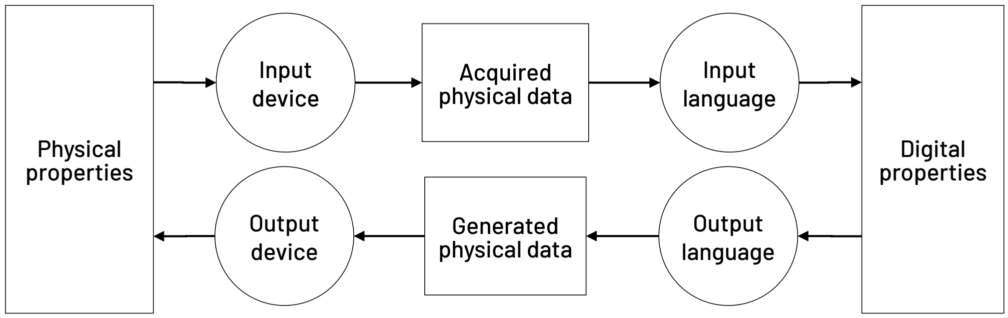
Before defining my own modes of interaction, it is helpful to discuss where the artefacts of this thesis are situated alongside existing frameworks and models. This provides a foundation upon which the proposed modes of interaction are built, and places the methodological contribution of this chapter within a broader theoretical context.

The theoretical investigation of human-computer interaction design begins by categorising interactions into three major paradigms: the *computer-as-tool* paradigm, which seeks to extend human capabilities with computer systems; the *computer-as-partner* paradigm, which gives computer systems agency to interact and 'collaborate' with the user; and the *computer-as-medium* paradigm, which focuses on computer systems as platforms for communication and expression (Beaudouin-Lafon, 2004). These paradigms that *categorise* interaction sit above the lower level, operational analyses that can be used directly by designers, such as *interaction models*, *design criteria*, and *frameworks* for interaction design.

At this level, Coutrix and Nigay (2006) define a model of *mixed reality interaction* with *mixed objects* — set apart from physical or digital objects. This model sees mixed objects transform physical inputs into digital outputs and vice versa, in processes Coutrix and Nigay define as *input linking modalities* and *output linking modalities* (shown in figure 7.1). Input linking involves capturing physical properties via a *device* and interpreting them into a digital environment through a *language*. Output linking refers to the generation of data from a language of digital properties, which are translated into perceivable physical properties through a device.

In the context of *live performance*, A. J. Johnston (2009) defines three modes of interaction with a series of virtual audio-visual instruments. *Instrumental interactions* provide continuous and predictable links between performer actions and virtual instrument response, ensuring a performer's control over the virtuality is

Figure 7.1. – *Coutrix and Nigay's* mixed object *model.*



consistent enough for nuanced performance. *Ornamental* interactions accentuate aspects of a participant's existing performance gestures, where the virtuality exists to augment their practice mostly as-is, and operation occurs without a requirement of direct attention from the performer. *Conversational* interactions give some level of agency to a virtuality to shift the balance of control between performer and instrument, and align more closely with the computer-as-partner paradigm. These modes of interaction are applied to describe and evaluate expressive performance systems.

While Coutrix and Nigay's mixed object model or A. J. Johnston's modes of interaction have wide-ranging applications in mixed reality and audio-visual HCI respectively, it is also common for researchers to present frameworks with a narrower set of applications. For instance, in the context of their system for geographically-distributed teleimmersion, Kurillo and Bajcsy (2013) propose three modes of interaction: *first-person mode*, *third-person mode* and *mirror mode*. In first-person mode, participants view their remote partner in the virtual scene, without seeing their own representation. Third-person mode lets users observe the scene from a perspective over a reprojected image of themselves (as in third-person video games). Mirror mode presents the user and remote participants front-on and horizontally-flipped, so to resemble a real-life mirror. This work presents highly-specific modes of interaction to offer insights within the single context of teleimmersive scene design.

Presenting these few existing models of interactive system design demonstrates how its theory is approached at various levels of abstraction. The modes of interaction I am about to introduce sit within the context of these works and extend their theory further. Firstly, these modes operate within the *computer-as-tool* and *computer-as-medium* paradigms. They conform to Coutrix and Nigay's mixed object interaction model, though are further delineated by exploring the types of physical properties that translate into the digital representation. They can be classified under A. J. Johnston's instrumental and ornamental interactions, and likewise enable the description and evaluation of expressive audio-visual performance systems. Additionally, they describe modes of mixed reality and teleimmersive presentation like Kurillo and Bajcsy, though instead of concentrating on modes of *viewing*, they shift focus towards the *placement* of virtual elements in mixed realities. The upcoming modes of interaction build upon this body of mixed reality system design theory, extending it with novel perspectives within the narrow context of live performance.

7.3. The four modes

Each system designed for the performances of the creative portfolio detailed in chapters 4 and 5 embodies distinct modes of interaction that enable real-time creative expression in mixed reality. To provide a comprehensive understanding of these artefacts, this section outlines four significant modes of interaction that underpin them: *introduced virtuality*, *augmented physicality*, *embodied virtuality* and *displaced physicality*.

Each of these modes represent a unique way in which physical and virtual elements can interact in a mixed reality performance environment. Some performances discussed employ more than one of these modes — simultaneously even — though each mode is discrete. They are defined by:

1. the input methods used to construct and control virtual elements, and

2. the output methods used to place virtual elements in a performance environment

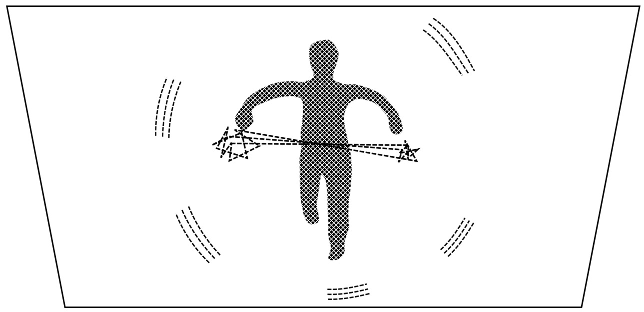
As such, each mode of interaction explores the question: how is this mixed reality grounded in the physical?

7.3.1. Introduced virtuality

The first mode — *introduced virtuality* — refers to the placement of totally virtual constructions into mixed reality environments. This interaction is defined by a participant's engagement with a mixed reality construct that is imagined by the designer and synthesised by the system. An introduced virtuality can certainly react to actions from participants — its behaviour can be tied to physical gestures and other interactions — though its basis for existence is not grounded in the physicalities of the environment it is placed in. It is by its nature *introduced* into the physical environment from the virtual realm. An introduced virtuality could be taken and placed among any other physicality — any other performer, or any other performance environment — and it would behave the same way.

A straightforward example of introduced virtuality is in the audio-visual instruments played by the AirSticks as presented in section 4.3. Shown figure in 7.2, the *constitution* of virtual instruments is unrelated to the physicality present on stage behind them. These constructs are highly-interactive — they are manipulated freely through the use of the AirSticks, and totally controlled by the physical gestures of the performer — but their existence in a mixed reality is in no way dependent on the physicality of the performance environment. Moving these virtual instruments to a different venue, or positioning them in front of a different performer will not change their virtual characteristics as their input is not sourced

Figure 7.2. – *The AirSticks' virtual audio-visual instruments are an example of* introduced virtuality.



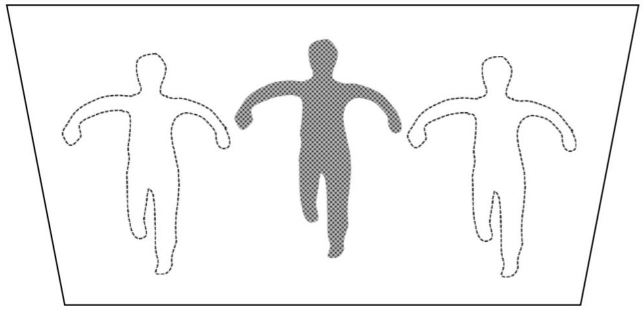
from either of these physicalities. Their digital forms are introduced into the physical environment of the stage for each performance in the same way a traditional, non-digital instrument would be.

Introduced virtualities represent a simple form of mixed reality object. Since the input modalities of such an object are not inherently grounded in any physical aspects of the mixed reality, particular care needs to be taken to effectively integrate the virtual construct into the mixed reality scene. A basic superimposition of a virtuality into the physical performance environment may not create a convincing or engaging mixed reality object. Therefore, in designing for this mode, it is important to consider how meaningful interaction with the introduced virtuality can be facilitated. This could involve designing a rich array of control gestures with a device like the AirSticks, or more simply just designing synthesised constructs so they respond clearly to performance gestures in a manner that feels intuitive and appears appropriately reactive. In the end, the goal is to design introduced virtualities that fuse into the physical performance environment despite their independent nature of construction.

7.3.2. Augmented physicality

The second mode of interaction explored within this collection — *augmented physicality* — focuses on the use of virtualities on the amplification of physically present

Figure 7.3. – Trigger Happy Visualised *extends the performer's presence on stage through* augmented physicality.



components of the performance environment. Augmented physicality develops from some aspect of the physical context it is presented in. This interaction mode presents a *virtual extension* of a physicality in a performance. It could be an extension of a performer, their gestures, their tools, or the performance environment itself. In contrast to introduced virtualities, the virtual components of an augmented physicality are inherently grounded in the physical world from their construction. Changing the physical aspect they are tied to will consequently alter the virtuality. This mode of interaction embodies a dynamic digital augmentation of the performance.

The production *Trigger Happy Visualised*, presented in section 4.2, contains multiple scenes that demonstrate this mode of interaction. In some scenes, the physical input modality of the performer's own body is used to construct virtual elements that work to amplify their performance gestures and presence on stage. One such scene places multiple real-time 'clones' of the performer onto the stage environment that mimic the physical gestures made — illustrated in figure 7.3. They are distorted and manipulated into expressive virtual forms beyond reality, though their shape is sourced from the performer's physicality in real-time. If another performer were to stand on stage in front of this scene, the virtuality presents itself differently for the new physical input form.

Augmented physicalities can also amplify inanimate parts of the physical performance environment — such as the drum-kit used in *Heifen* (discussed in section 4.2.3), that is augmented with bursts of colours when it is struck. In this example, design of the virtual construct is *spatially aware* within the context of what's on stage. Unlike the AirSticks virtual instruments — established as introduced virtualities — this mixed reality interaction is foundationally connected with the physicality of the drum-kit. If the drums are repositioned on stage, the virtual augmentations shift with them, and if they are totally absent, the virtual construct ceases to exist.

The mode of augmented physicality therefore, presents a virtuality that is inherently more embedded inside the physical. This grounding makes the integration of such mixed reality constructs easier for participants to comprehend. As long as the connection to the source physicality is well represented, there is less work on the practitioner's part to explicitly integrate the virtual object into the mixed reality environment. The object is already a mixed object as soon as it is constructed.

The captivating function of augmented physicality is its exploration of the equilibrium between the physical and the virtual in a mixed reality. The goal when designing for this mode is to strike an effective balance. A virtuality that overpowers or distracts from the physical source in this mode defeats the purpose of its grounding in physicality. Likewise, if the augmentation diverges too much from the physicality, is it still an augmentation? Core to the interaction mode of augmented physicality — and its main challenge — is the creation of a virtual object that is both a coherent extension of its physical origin, and a distinct entity that offers new capacities for creativity in performance.

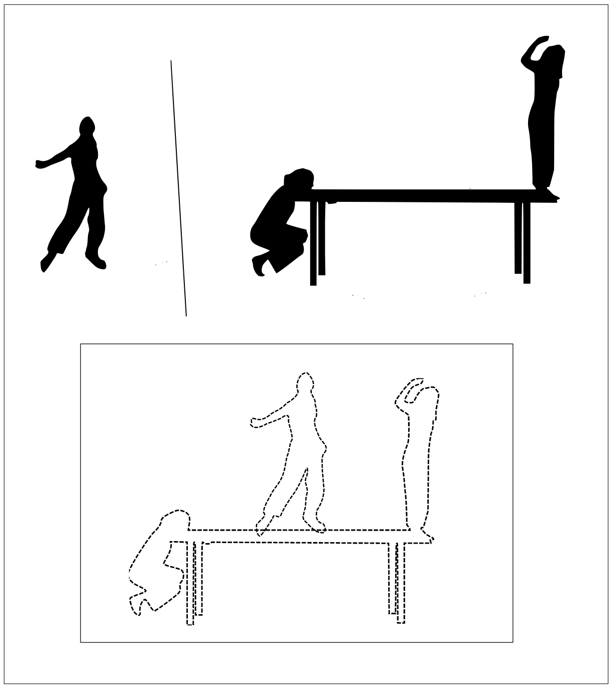
7.3.3. Embodied virtuality

The previous two modes of interaction explore methods that display virtualities superimposed into a physical performance environment. *Embodied virtuality*, on the other hand, is not concerned with an output modality that is integrated within the physical environment. In this mode, the participant uses their physicality to embody constructs inside a separate virtual environment, where presentation occurs detached from the physical environment itself — not superimposed among it, and not used to augment it. Participants engage with embodied virtuality by observing a separate screen or projection, which displays a viewport into the virtual environment that contains a reconstructed representation of their physicality. Participants interact with other entities within the virtual scene through the movements of their virtual embodiments. This mode of interaction extends the concept of teleimmersion within the context of live performance.

The mode of embodied virtuality was explored through several of the productions presented in the creative portfolio. In *Computer Storm*, the participants — both performer and audience — are virtually reconstructed and placed within a distinct virtual environment projected onto a scrim. Even though the scrim is located within the performance area, this presentation medium itself does not contribute to the construction of the mixed reality. It is not used to embed virtualities into the performance area. Instead it displays the separate virtual environment, initially constructed by the physicality of the participants. The users embody their virtual reconstructions, which are further manipulated by the performer's gestures to control virtual elements transforming the scene.

Additionally, the geographically-disparate mixed reality performances presented in this thesis such as *AirStream* (5.2), *Layer* (5.3), and *My House Your House* (5.3.6) all make use of embodied virtualities. Shown for *Layer* in figure 7.4, each production captures remote participants and places them together into virtual environments as a means of facilitating internet-connected performance. The tertiary, virtual

Figure 7.4. – *Multiple scenes in* Layer *enable embodied virtuality. The Groningen table scene places virtualised versions of all three performers inside a virtual environment that is projected seperate from the physical input environments.*



performance environment contains the constructs that become the central focus of the performance for all participants. In *AirStream*, the performers themselves aren't visible to the audience — only the embodied virtual constructs — and in *Layer* and *My House Your House*, the audience see only one of the physical environments with the other being distant. In all cases, a shift of focus away from the source performance area and towards the virtual mixed reality environment occurs.

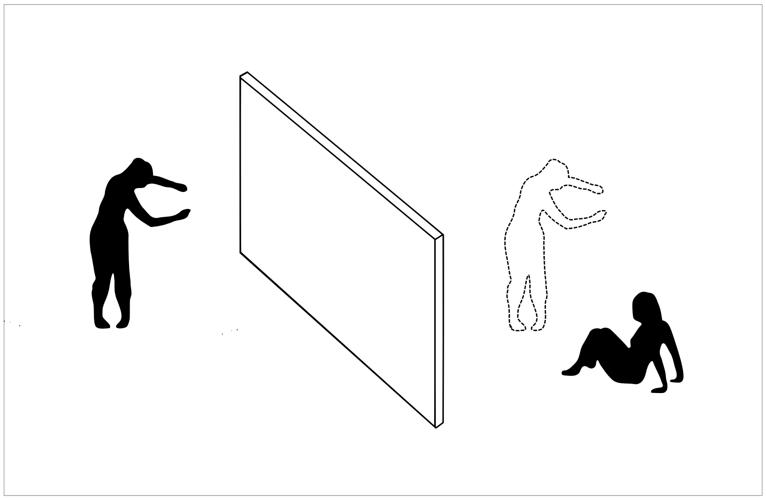
The utility of this mode is in its ability to provide natural interaction within the virtualised setting. The embodiment of each participant's physical form within the virtual environment allows for intuitive action and reaction as they watch their reconstructed selves engage with other embodied virtual partners, as well as synthesised virtual elements. The shift in gaze required to properly engage with the virtual environment can however be restraining. In *Layer*, the performers developed an instinctual idea of where they may be positioned inside the virtual environment, though initially the dancers strained their necks to keep their digital selves in sight, and needed to be guided by off-stage direction. Multiple monitors and projections of the embodied virtual space can be placed around the physical area so participants can see themselves and their partners, but the fact remains that any connection created through embodied virtualities must be achieved though this physically-distanced view into the mixed reality. Unlike other modes of mixed reality interaction where virtualities are situated in a position inside the physical performance environment, embodied virtuality requires an additional layer of practical and conceptual consideration. Interaction is not simply an extension of the physical space, but is mediated through a tertiary, separate, virtual environment.

7.3.4. Displaced physicality

The final mode of interaction — *displaced physicality* — describes a modality whereby a participant's presence is transferred from one physical performance environment to another through mixed reality. This mode involves first the capture of a performer's physicality into the virtual domain, then the reintegration of the resulting virtuality back into the physical world. In displaced physicality, virtuality serves as both a conduit and a canvas: the virtual environment is a transient medium that facilitates the relocation of one physicality into another, while also offering potential for transformative, creative reinterpretation.

Displaced physicality was explored in both *Layer* (5.3) and *Critical Path* (5.4), where distant participants were integrated into a remote physical performance area. In the case of *Layer*, the performer in the Netherlands was embedded onto the stage in Sydney by way of 2D projection. In *Critical Path* (5.4), performers were shifted from an off-stage location to on-stage with other performers — as illustrated in figure 7.5.

Figure 7.5. – Critical Path *showcases displaced physicality through its transferrence of a remote performer into a distant physical environment.*



In contrast to embodied virtuality, where connection between participants and virtual constructs are established through an abstracted virtual environment presented in a separate viewport, displaced physicality allows these connections to form within the performance environment itself. The virtual reconstruction of a participant's presence coexists with the in-situ physicality of the stage environment for a higher level of integration between the virtual and the physical.

7.4. Conclusion

The modes of interaction presented in this chapter — *introduced virtuality*, *augmented physicality*, *embodied virtuality* and *displaced physicality* — do not attempt to encompass all possible configurations of mixed reality performance. In fact, they articulate the opposite. Derived from my specific experiences developing the creative portfolio of this thesis, these four modes of interaction serve as a lens to help explain key concepts that emerged through development of new live performance with mixed reality technology.

The distinct relationships described between physical and virtual span the width of the virtuality continuum, and are presented with the intention of broadening our understanding and providing guidance on what mixed reality is able to offer to live performance. This theoretical contribution can be used to frame the description, evaluation, and ideation of future mixed reality performances.

8. Conclusion

This thesis has explored the integration of mixed reality technologies in live performance. A practice-based exploration was conducted that centred on the creation and presentation of several new productions which fused the physicalities of live music and dance performance with the possibilities of integrated virtualities. The seven performance works presented as this thesis' creative portfolio throughout chapters 4 and 5 form a body of work that highlights a variety of opportunities and challenges presented by the burgeoning technologies of mixed reality. Across these productions, the task of designing systems to accommodate the multifaceted needs of artists, audiences, venues and the technology itself was consistently at the forefront. The variety in scope explored within this collection of work is testament to the vast possibilities that mixed reality technologies have to offer to the world of performing arts. As the expanding mediums offered to us through the development of mixed reality become increasingly ubiquitous, their potential as a canvas for creative expression will only continue to grow. This thesis aims to promote this growth and aid fellow researchers and practitioners in their ambitions to do the same.

The three objectives of the research (stated in 1.2) were:

1. Investigate the potential of mixed reality technology in music and dance performances through a series of experimental productions

2. Gain insight through these productions into the challenges, successes and opportunities presented when integrating mixed reality with live performance
3. Develop and refine a performance-focused mixed reality system based on these discoveries

The following key contributions of the thesis realised these objectives:

1. A creative portfolio of novel mixed reality performances

Significant creative artefacts were presented as the primary output of this thesis in chapters 4 and 5. These works serve as demonstrations of the artistic possibilities offered by mixed reality technologies, highlighting new and meaningful approaches to the construction of technologically-mediated music and dance performance. This collection of creative research artefacts span a wide range on Milgram and Kishino's virtuality continuum. The concepts of augmented reality, augmented virtuality, teleimmersion, and telepresence are all represented — and the application of several modes of interaction between virtualities and physicalities within diverse performance configurations offer valuable discourse and direction for future production in this field.

2. Performance-focused mixed reality system implementations

The underlying designs of the systems that enabled the mixed reality performances of this thesis were discussed in their respective sections within chapters 4 and 5. The elaboration on the practical implementations of these systems provides useful strategies and considerations for practitioners seeking to implement similar productions. The synthesis of this knowledge was presented in chapter 6 — first as an exploration into key learnings relating to software and hardware implications, and second as the presentation of the new system for mixed reality telepresence, *pointcaster*, designed specifically for live performance with augmented reality headsets. These outputs — grounded in practice — offer valuable insight on mixed reality system design in the context of live performance, and provide guidance on technical considerations in the areas of depth sensor technology, point-clouds, mixed reality presentation methods, networked performance, interaction design and more.

3. A framework for the development of mixed reality performances

Theoretical artefacts were presented in the form of a framework for evaluating and guiding the development of mixed reality performance works. This framework was split into two components: a set of *system design heuristics* relating to the technical development of systems for mixed reality performance, and a collection of *modes of interaction* that guide discussion and development of such systems through a purely theoretical lens. The system design heuristics, presented in section 6.3, emphasise four critical criteria that promote successful system implementation: *adaptability*, *resilience*, *expandability*, and *creative control*. The modes of interaction, presented in chapter 7, identify four distinct composition models of virtual and physical entities in mixed reality performance: *introduced virtuality*, *augmented physicality*, *embodied virtuality* and *displaced physicality*. This framework provides approachable and structured methods for the ideation, design, and analysis of mixed reality performance — positioning itself as a resource for researchers and practitioners looking to push the boundaries of this medium.

It is my hope that the contributions of this thesis ultimately provide inspiration for others interested in the integration of mixed reality into live performance practice. This domain is fascinating, and the continued growth of mixed reality technologies is visibly turning what was once limited to the pages of science-fiction into a reality that literally surrounds us. The world is primed for virtual, screen-based media to break free of its two-dimensional confines and embed itself in our physical environment. I cannot wait to see how practitioners will push this medium into the next generation.

A. Interview

Initial discussion

Me: can you tell me about your experience at the residency?

Choreographer: I think what was good is that we kind of had a through line but at the same time we didn't have any expectations, which felt great in terms of being able to really research and figure out what it was that we wanted to do.

I think we had these parameters which were really important for me about the ghost dancer which is something that I think we kept all the way through and I was really happy about that.

It was a very different — how can I explain this — it was a real gymnastic of my brain to figure out ... my right from my left when I was in the other space ... in the glasses space.

This confusing thing about the real, the augmented reality and kind of the virtual as well because I think there's still the virtual too within that even though it's not you know virtual [reality] headset I guess.

So really loved that and also got I guess sometimes frustrated with it because "what do you do with it?".

I think there's a huge potential with the glasses to really push the technology but push choreographic writing or how you anticipate movement or how you develop movement.

Yeah, how you construct a piece as well because it's already hard enough to construct a piece in one space, but when you also have to think about the augmented reality space … I had to think about what people were seeing with and without the glasses.

I think for me it's that interplay between the real and the augmented and the virtual and you know if we were to continue with this it would not be just a show for people wearing the glasses — I think there's also a show for the people who are not wearing the glasses, and that is something that I got from the feedback — especially from the less technology people — more the dance people — they've really liked watching both and see how it affected perhaps the narrative, or how they think or their critical mind.

It was really useful to have another body there like Imogen because it takes so long.

I mean in choreography to create you know 30 seconds it may take days but with add that with the technology I feel like you need like three months to make 20 minutes really because it's, yeah, it's a slower process because our brain is not yet wired to think that way, maybe or at least mine.

Perhaps maybe in the future we develop an extra little bit in our brain I don't know… or you know the brain rewires itself to to think yeah I don't know it can be it's confusing and at the same time it makes a lot of sense too…

I really also like that you as the tech guy — let's call you the tech artist, I prefer that — you really became involved in the choreographic making, and also because I think you used to now also work with choreography so you understand it to an extent just like I understand the technology to an extent. Very limited, but it's important. I think it's this interplay between the two it's a real collaboration. You need that I think you need someone like you can understand the choreography or the spacing. I think it was a good collaboration in that sense.

M: jumping on that, you have worked with other collaborators before in different technologies right? was the relationship between choreography and technology different to the other times you've choreographed using different technologies?

C: I think that as soon as this technology there isn't there's a pressure to make the technology you know 'Wow', and like, the focus goes whether you're conscious of it or unconscious of it — the focus goes to the technology somehow.

I think maybe because it's still fairly new and there's still so much to discover with it and it's almost like suddenly the choreography becomes a safe bet. Or enough to complete the digital. I'm not saying all the time, but I feel sometimes we think too much about how high is it gonna look with the digital instead of thinking about how can the digital support the choreography, we're seeing it as how can the choreography support the digital.

Because I am a choreographer and sometimes I just want to concentrate on the dance. Sometimes with the digital set — how can we make it look nice — I'm really simplifying here — it needs to look nice well then I'm interested in more maybe the ugliness or the beauty of the ugly or making something that is not as beautiful shapes and more emotional.

So I think in the future maybe it's like how can the digital complete that, because I feel sometimes with the digital it's so far removed from it and it's just "look nice".

There's also something for me very scary about the digital taking over.

For this — and also because of the parameters of the residency — it needed to come we have the choreography or we develop the choreography and the digital is there as a response or completing, but it's not like the technology does "this" so we're gonna do "that" choreographically.

I wanted to go the other way around to see if it's possible. Maybe it was gonna fail but I don't think it failed. And if anything, now we can find the balance between the two a bit more because there's that as well.

Sometimes I feel like the digital takes over and that's you know the fault of our phones we're always on it and with such a visual... we like colors as well I think that I attracted to colors so sometimes maybe we forget the dance and you just look the same when you have the dancers and the digital.

But with the glasses actually I think it's less like that. You're more immersed, so it's more like watching a live performance. the digital is within. It's better than the screen definitely, or a projection. There's something that is more immersive in it, which I really I really enjoyed.

You can and you can feel the presence as well, which was another thing that I really like. And when you get used to it and you understand how it works the right the left and the directions, and you remove the glasses but you know your choreography it's funny how you can feel the hologram in a way. You can sense this augmented partner that you can't see when you remove the glasses.

Once you've established and worked with the glasses, you know where the augmented body is, and then you remove the glasses and you repeat the same thing. You have that connection, that ghost, which is incredible...

when I was dancing with Immie and she was behind a mirror, I couldn't see her but I could sense her so we had an obstacle there between two real bodies.

And then the obstacle is like, not having the glasses between the real and the projected image. And it's like "oh I can feel that too".

M: Well you've kind of started touching on another question I had, which was do you have any thoughts on how the three dimensions of this technology compares to some of the two-dimensional stuff that we've done before, like Layer and MiCasa. Where those were projects in a 2d space. How did it change your experience in terms of as a dancer, as a choreographer, just being in 3d space now?

C: I think for me, when I make something without technology there's always that audience participation that I seek because I don't particularly like making something for people to just watch. I like them to feel involved in the piece so it's like breaking this fourth wall. I think the glasses to a certain extent do that without the performer having to do too much. Really it's more about you as a programmer like you create that breakage of the fourth wall, which aesthetically or narratively, creatively, is kind of my taste.

Also there's this idea when it's 2d and it's flat, it's all linear so the choreography becomes linear. With the glasses it becomes circular — or like I like to imagine it like a little globe like the earth — it's just you, surrounded.

You know, in the future we touched on, having furniture and augmenting that furniture, and I think there's something we can push there in terms of the set design. Like in the architecture, we've talked about that so I think there is a huge potential there again which you wouldn't really get the same way in a live performance it would be flat or in 2d.

I think with the glasses there is just this illusion that you can grab. The illusion that your body is actually involved.

I don't know ... I should research that — or maybe look at phenomenology... It's like it's missing for dance as well, and you know when you see a body, there's this link, but it's created emotionally and there's all this phenomenon. That's kind of what I seek when I make, and I think with the glasses there's a different phenomenology happening.

I don't know if anyone has written about it actually it would be interesting...

I think it's definitely in there with the glasses. It's your body. You become like a child. You just want to touch it — everything — and you know you can't touch it but, somehow your brain doesn't quite comprehend it.

It has a magic, it's very magical. What I like about these [AR] glasses as opposed to a virtual reality headset is you don't get lost as much. Reality is still present, which I think philosophically you can push so much ... A virtual reality scares me. I just feel like you can really get lost and never come back in that one. Within the [AR] glasses, I'm still present, and I like that sensation.

... the 2d — now that I've tried the glasses — feels a bit obsolete to a certain extent, and I don't really want to go back to that. Just having flat projection is nice, but I think now it's not enough. Maybe for where we're at.

M: You said you get that presence of Imogen as a dancer, performing the choreography, and I just wanted to ask, did you ever get any kind of similar feeling when you were working with the team in the Netherlands? Was there ever any kind of presence as a dancer? could you feel Olympia? could you feel where she was in the space or anything similar?

C: I guess yes — because for a very long time as well the other physical dancer I was working with, Katina, wasn't present. So I think I developed this dependence with Olympia, because we're trying to create a duet. I think the presence was more the image. It was a creative, the connection with an image was always someone that was not me.

With the glasses, we also tried another image that was me. So again, you know, philosophically it's a weird thing. It's like an alter ego there or another bit of you. So inherently you're always connected to that, but then it's projected, so it's a weird thing. It's like a mirror, but not quite.

So with Olympia, I realise doing the performances, of course we were connected, we had to be together, but I wouldn't say it was a magical experience. Sorry to use this word, maybe it's a bit silly, but I really think it's like magic.

With the [AR] glasses I could feel the presence through the layers of the skin. There was a tactile sense. Where with Olympia it was more my brain was connected. So and I knew my movement of course, we had connection, and I connected to her image. My brain knew that she was in the opposite side of the world, so I was like that's where the magic was. We are dancing together but you're somewhere so far away although we can now do that with the glasses. I'm sure we could. Maybe that's the next step to see what is the difference, but yeah the connection is not as strong as with the glasses. It really feels like it's another body in the space.

I also think like the level of precision with the 3d glasses — I mean I know you had to be very precise knowing that it could never exactly be the same to well you know millimetre of the image that was projected — but what I loved is that, you know, if we rehearsed it with my arm being like this, I had to really know in my space where I needed to put my arm again and again and again. And of course in dance we have that, but I think it was amplified because of the glasses.

With the 2d you can be a bit less precise and it still works. There's still a sense, but it's more about I the proximity or — how can I explain it — it's not about the subtleties with a 2d it's more about the general look. With 3d glasses it's all about the subtleties because you can make it really fit well. With 2d the audience won't see the difference really. It's more "okay the bodies are here", it's more about the angles, so in terms of a choreographic tool I think the [AR] glasses are a hundred times better — and a hundred times more difficult sure — but that's the beauty about it. The precision and attention to details.

M: Did the technology allow you to explore anything novel choreographically?

C: Yes of course because working with glasses in itself is already a new thing. It's me thinking how to guide the dancer. It's how I guide her spatially and have to work in absolute real-time with her.

Also you know the frustration, I guess one of the negatives — but also I still think it's a positive in many ways — is this precision that we had to have. Like the fact that sometimes we really need to use that tape to almost like a cameraman with their angles is very much more like a cinematography way of working in a sense because as a dancer… we don't use tapes that much — like for centre, and there might be one or two for one hour show and the less tape you've got the better. But for this one there was quite a lot of tape on the floor and I really enjoyed that too.

We had the front and the back, and the back was the tech side and it's also I think really for me acknowledging it's technical — it's technical for you, it's technical for us, it's also a bit technical for the audience.

I think the dream is also to choreograph the audience. That's also something that the glasses give is anticipating what you want the audience to see. So we could have really also choreographed them even more, but that takes so long. I think that's a dream. I love doing that, because I like manipulating the audience brain and I think with that I can really manipulate them physically, without being too heavy on telling them what to do.

I wonder if there's a way to to make it so smoothly that they don't realise that they've been manipulated. I think that's what technology does to us, it manipulates us, so yeah I think as a narrative or a thread for further research, that's really something to explore. The power of those glasses is huge.

M: What are you trying to get them to experience?

C: So, one of my interest is the gaze. Manipulating the gaze. It's as simple as that. You know, we do that in a live show. You do that with lighting, you manipulate where you want the focus to be, but it's just the eyes that are working. I think with the glasses there's a potential of where their body is in space instead of having them static. If you have them static then suddenly you go back into that fourth wall where the audience is just an audience. I think for me technology is about interaction, but also art is about interaction.

I don't like a lazy audience and with technology you can become lazy, so it's fighting against what the technology is or can and can't be. So you make them active, but still keeping that. I think the manipulation is maybe one of my vices. I just I don't like it as a human being, but as a creator… I think if you write you manipulate the reader. As a creative you manipulate, let's face it. You want them to see a particular thing even if you don't tell them when you want them to react. You want to provoke. I think I like provoking. So my provocation, because there's voyeurism in the glasses, the way we worked was very intimate we know what we get ourselves into. It's a technology world but there's two human bodies in there. It was two female. So that has an impact on the narrative even if you don't set the narrative yourself, and tell them this is what it is about. The audience, you know with the phenomenology, is going to react in a certain way just to that setting. Then you put the glasses on top of that and you can just create your own narrative.

As a maker I want them to see that beat because then I can jump from what I think they may feel in that moment. We're not all that unique. At some point, our emotions just happen. You know, if you give us an image, we all react pretty much the same way. I think more or less, so yeah I think it's like "how far can we manipulate an audience without them realizing it?" To show the danger of technology. That's something that would interest me.

M: Could you elaborate on some some of the audience reactions that you noticed? Did your provocations work? Did they react in unexpected ways or expected ways?

I loved on the Saturday one of your friends I forgot his name. He was wearing white pants and very fancy Converse. Elie? Yes! I really liked how he wanted to discover the back — the other side of the mirror — because that was something I really wanted to put in, but we didn't manage.

I think we pushed back on how much we wanted to manipulate the audience for that particular showing as well but seeing him going — and Julia as well a bit — exploring that back it's like the hidden side, the forbidden side maybe, or the thing we often try to hide. I guess in my mind I started to imagine "okay what can we do with that?", like by looking at what the audience wants. In a way it's not really me that manipulated them, in a way they manipulated me, or manipulated, you know, the idea where could we go with it.

.. I guess it will be for a next step. How we can manipulate the audience to really want to see what's behind but maybe forbid them from going there until we're ready to show what we want to show at the back.

"What is it that is so special about that back space?" I realized the people who work with technology had a different reaction. There was another person ... I really liked that he wanted to be closer, which I think to a certain extent we talked about, and I think David expected people to go closer to the holograms, but in my experience I feel like people tend to be polite especially in that setting they stay a bit more back.

So again, we had a lot of failed manipulation of the audience I guess which we can learn from. And it was okay, so next stage: how can we get people to go closer? how can we force them to go closer without it feeling like they're forced? There's something that's in those images from close-up that is amazing. I mean the manipulation was at the beginning what I wanted them to see as opposed to the expectation of you know I'll put the glasses on now I'm gonna see something that I haven't seen before. It's like no! You're expecting that so I'm not gonna give you that! That was the first manipulation I think and then telling them now you can move a bit the other manipulation.

On Saturday when we had too many people — which I thought was great because then it was "well you're gonna wear the glasses", "you're not gonna wear the glasses", "you can do whatever you want", but we manipulated the experience.

I don't know if I can say more about this because I think in a way the manipulation failed in many on many levels but because we didn't spend that much time on it, it was kind of like an accidental manipulation to draw from for next time. It's just so so much to consider with a work like that.

I don't think don't think I've got more on this.

M: That's fine. I think we've kind of gone through a few of these questions that I wanted to ask. Okay well maybe we just look through some videos now.

C: Okay

Video prompts

M: So we'll watch the video and then I'm just gonna ask you to explain what's happening or elaborate on anything that comes up. Also if we're in the middle of the video and you feel like saying something we can pause it.

C: (in reference to watching herself in the mirror of the video)...That was funny. I'd forgot I was wearing the glasses. I think at some point you can tell because when you choreograph you look with different angles and I can see this clearly what I was looking for.

So I guess going back towards the manipulation of the audience — that's exactly what I was doing. I think when I was making it, I was trying to manipulate them — creating that illusion — going through the mirror or being two bodies or leaning against. So actually it was not such a failed manipulation, because a lot of things happen and succeeded.

I really remember actually when we filmed this now. So there was the mirror and I wanted this hand to look like it was on the mirror. It really looked like that the first time we did it — that she was placing her hand on the mirror from my corner. So [here] I get really obsessed about placing the audience in the right angle, which is impossible because you have a perimeter that can make them see that moment. Which is really hard to to direct.

When you make choreography you just make what you like, but with the glasses the whole time I was really thinking about the audience. I realized it was all about the audience.

M: Is it not what you usually would do?

C: Well, I always think about the audience, and you always have to think about the audience, but I think in this development it was really more highlighted. A show can still exist without an audience I think, or even if you have just one person, but with the glasses you really need to have them in a particular point otherwise it becomes completely flat and there's no point them wearing the glasses.

So I think that's where we discovered that the diagonals were so important. So when you work in 2Ds diagonal is not good but in a 3d context diagonals are so important... I mean of course flat from the front works too, but I think there's a very good hot spots in the diagonals. If you work with an obstacle in the middle — like we had the mirrors in our case — I think that creates very nice imagery. I think it's really much a work of imagery with the glasses because it has that filmic component.

Also it's nice to watch Imogen trying to orientate herself as well. You know precise we had to be, when we were pressing these buttons as well, and for her to understand where she is in space. She's used to working with technology and also is a tech person I guess, but it took her a while to understand this. Probably faster than any other dancer — she was the perfect person for this.

Then you can hear me guiding her ”go back”, like I think I knew what I wanted to see or I let it happen in real time, so I let my own desires of this is the illusion I want to see. So it’s subjective but as I said, I don’t think we’re that unique, so I’m sure that people think ”oh wow it looks like she’s going through the mirror”, ”oh she looks like she’s behind the mirror”.

I know that behind the mirror I was bit annoying with that because I really knew where I wanted her exactly to give that illusion. And the millimeter forward it was just not the same effect.

I guess now a wish for the future ... with a hand for example going through the mirror, you had the frame of the mirror and you could see the the arm on top of that frame. So I guess the wish for the future is that you really create that illusion that she’s going through. You have the frame at the front instead of at the back ...

M: when first the video opens, you’re making some remarks about the clothes...

C: Yeah, I really I remember I loved the particular black t-shirt. It didn’t work with all black t-shirts, but the one you were wearing disappeared, so it was just your head and your legs ...

I wanted to look at that and just having body parts segmenting the body to create abstracted shapes ... so the real was dictating the digital in a way again. Instead of the digital doing its work, it was coming from the choreography imposing on the digital, if that makes sense ... removing body parts because of the clothing, I really wanted to do that ...

M: We haven’t really talked about the phone and physical constraints. Immie wasn’t wearing the glasses in the performance but she is in this video. Can you elaborate a little more on any of the physical constraints imposed by this tech?

C: Well to develop the work we had to both (choreographer and performer) wear the glasses ... to know what she was doing roughly, to get a sense of what was the image that we were creating. The limitation was that the headset, well the phone itself was getting so hot. Also to attach it to the body ... we put it in the pocket and we clipped it so it wouldn’t fall, but it was getting too hot in the pants so we had to take breaks. The headsets themselves were getting hot too.

I think for Immie to really move freely and not look like "oh, it's a work of technology". Those glasses give a Matrix look, so it gives a narrative to an audience and I didn't want that type of narrative. So that's why we developed the choreography that way. For her to know exactly in space where she needed to be, so then she could remove the glasses but it embedded within. She would program her body to something very accurate. I mean I think that's why there's that link between the projected image and the actual body. You can almost feel it it's because you've worked with it, and it's embedded in your brain. You can sense where it is.

I remember with the hands, I could really sense the hands up there and I felt I was really playing with the hands, but I didn't know in real time what it was looking like, I could only imagine. It's a scored imagination. We have scored choreography, scored improv, and I think this is a scored imagination ... or scored augmented reality I don't know, we should name that actually.

You can't stay in headset for too long, neither for the audience nor for the for the dancers because you fatigue. I mean we had to get used to it ... by the end of it we could actually stay in it for much longer than we could at the beginning, but your eyes get a bit tired.

I didn't move that much during the residency but I think the thought process was intense — more so than in other developments. You always think a lot when you're making, but with this you have to "double make" in a way. You had the glasses on, so you think about two realities — three if you include the audience.

We were pretty much doing nine-to-four, but I thought by four o'clock I was a bit done.

M: Do you think it was because of the way the glasses made you physically feel? it was a draining experience?

C: Yeah, it was draining — in a good way — because we didn't know what we were doing. I mean we still kind of don't know. We know something interesting is happening. I mean it's always great when you start a choreographic process. It's so freeing when we don't know where we're going, but we're gonna try that.

I think the glasses create expectations, especially in the audience — more than if there was not any glasses. There's the technology and people expect to see certain things through those glasses. I totally think the audience expectation is enhanced.

For me, when I was wearing the glasses, I had expectations of Imogen's hand being a certain way, and when it wasn't I was so frustrated ... Trying to repeat the experience was frustrating because it was never quite happening the same again.

I think when we developed the three holograms later on, that was also an act of precision and it was such a beautiful image and we managed to make it When you wanted to create that type of precision you also had to have a looseness to know that it will never exactly be the same. It was more about giving the dancer a range. So this is the area you can operate in and if you operate within this, something will happen. We knew that. So it was a part of chance, and sometimes it worked better than others. But again it also depends on where your audience was standing.

So I think frustration is kind of part of the development. ... although not for the audience. I guess the audience might not know what the other person is seeing. So it's like you win you lose, and I really like that. It's an individual experience in a way and everyone has a slightly have a different experience. I said we're not that unique but we still have a slightly different experience, and that's the richness of it. So more than a show where you just sit and watch, where pretty much everyone has a similar experience, I think with the glasses there is that "you choose your own adventure" a bit more. Which is empowering for the audience. It can be a manipulation and an empowerment, it depends how you want to see it. Maybe you can trick them into thinking that they're empowering themselves.

... I think for me the magic for that image is when you had the three holograms there touching, they were interacting, but because it was a frozen image as well, it was kind of leaving, departing from the virtual. And because of the lighting, Imogen looked like she was not quite real, not quite virtual but in this in between state. And then when she was coming back into it to morph into something else. So it was that interplay — becoming one and becoming separate — leaving behind an imprint in a way. I love this idea of imprints in general, leaving a trace ... the ghost again ... I think we were really playing with the ghost there, but who who was the ghost. At first when we entered this in my mind I always thought that the digital would be the ghost, but once we did this, I realized it changes depending on the framing of the image. The real body can definitely become the ghost and that opens up whole sets of different narratives. I really think it's a work of on philosophy I'd need to do more research around the image, the double, the ghost, but I think there's a lot in. It's a whole philosophical paper about it. We never really had just the virtual. There was always a physical that we had ...

M: I think depending on where you're standing when she's creating the sculpture might feel like just virtual, but you obviously have a sense that [Immie is] there anyway.

It's not like the source was ever not present in the space. Not like if we were to combine this with Layer they were actually in a different physical space.

C: ... the virtual is kind of doing a duet with them, even if you don't have the glasses, it's interplay between the two. With that through the ground image I think I push this one. I really wanted to separate the audience from what was happening in the real. For them to ignore it. I think some people thought, on the Saturday, the people who were not wearing the glasses but witnessing that scene, it was actually a very strong scene for them because they could see the body ... a bit in distress, and then the people with the glasses ignoring that body because there was something happening on the ground. That's when the digital took over for them. They didn't care about the real body. But for the person who was watching this scene [without glasses] it was very impactful. They really felt that the people consumed with the digital didn't care [about the physical], and there was something a bit strange ... so that's why it was great in that Saturday session ... That feedback was really important ... So I think we need two audiences. Two distinct audiences because it creates another layer.

For the people watching, they did say they could still feel the other body behind. It was interesting to them to really feel they had that connection. We established that connection with a physical. Because you wear the glasses you're just drawn to the technology, so that's what took over, but it's nice for the technology to take over in a seamless way without the dance, but it was still generated by the dance. And because you separated the gap between the physical and virtual such that they need to stay just with the image ... I think we it was quite rich. I think we had a lot of different frames and possibilities.

C: (talking about video instructing Immie) ... I wanted her to kind of like slide ... being tactile, how can you bring in tactility?

Often [this image] goes through [the mirror]. I loved that. I wanted to do more of that, but it's so hard to recreate, and again it depends on where your viewer is. So in a way if this happened exactly during a showing, it was accidental. I mean I had it in my mind this was a possibility, but I had to stay frustrated and let go of the wish for them to be precise.

... Maybe in the future you could put a filter to make it look like she's stepping through the mirror ... keeping them inside the frame and then she get's through.

I think with the technology, you don't want to move too fast. You don't want to do things that are too complicated, because less is more. It needs to be clear, and at least for now you can't do too much, otherwise it gets lost — the meaning.

M: What do you mean too much?

C: Well, in terms of the movement ... What I'm interested in with the glasses is not so much just a dance work. It's a choreographic work, which is different for me. Dance would be more putting dance steps in there, while this is a movement piece. A true exploration of choreography. Choreography of the space, choreography of the technology, choreography of the movement. It's much more complete in a way. You have to think about everything.

M: Is that were you expecting when you came in?

C: I knew it was I knew was going to be tricky. I knew about the trickiness of it, but it really expanded my mind. More so than anything else. I think now if I go back into a normal, without glasses, choreographic process, I consider the space differently. Looking from all the angles ... it developed me as a choreographer.

As a tool it makes you understand a lot of things about your habits, or your maybe your choreographic signature. You're much more aware. Sometimes with choreography if something doesn't work, you don't know straight away, but with this I feel like you know straight away ... There's a certain amount of editing you're doing in choreography ... but I like to take my time. Sometimes you just go "it's too soon to cut", but for this I felt like I was much less attached to things. It was more like scientific work, like "this idea is a fail", or "this is going to take too long, we don't have the time". I had to be more pragmatic in a way and more efficient. In the end I the showing was a 15-minute piece. 15 minutes took two weeks using the technology. It's actually quite a lot of work and it's quite complex, so at first I was like "oh, it's not going fast enough, we're not gonna make much", and then I realized how much we actually did. It's a lot.

Through our discussions you were telling me "no, I can't do that". It was great because in choreography sometimes we talk for hours and we waste time. At least you were like "Cloe just forget about it", and it was good to put back on track, but I think it grew me I grew as a choreographer through this definitely.

It's very different from other things I've made too, but not so much. I've always wanted to work with mirrors — it's a good excuse to work with a mirror — but also this double image. I always have a bit of a double thing in the works. I understand my patterns. I think through this it's just like "oh yeah, that's how I see things", "that's how I look", and looking at the film through the glasse also makes me now understand his is how I look at things ...

M: Let's move onto the next video. The next one is a little bit longer so if you feel like saying something, just hit pause.

C: How it started here was just the particles and then it transformed into the body. I find that so beautiful.

What I particularly like here is the hologram really looks like it's her ... I think for the future, depending on how the technology grows, we could push that illusion even further ... shifting between the reality and artwork. ... because of the mirrors we were playing with a segmentation of the body. I think this idea came up because sometimes [the camera image] is segmented. Like what I saw with your t-shirt ... I think if we were to do a list of things to explore, segmentation would definitely be important for me to look at. That's something that the digital gave to me. It inspired. It's not just me putting out ideas, but it's like the digital giving me a feedback to grow the idea. So it's a good tool.

M: Are you talking about using the bounds of the cameras? Is this what you mean by segmentation and playing with the bounds is what inspired that initially?

C: Maybe it was in the very beginning, sometimes we were out of frame or in one camera but not in the other, and you can see just one bit. So you appear and you disappear ... as opposed to having the full complete body, it's just the face ...

it's very linear or maybe an angular idea of the choreography, or the frame ... Doing these things which I think sometimes happen when we were just workshopping — not so much in this final thing. I think in a way we lost the segmentation within the hologram but we put it in here within the framing as an echo of what I had experienced through the glasses ...

I use the music at the beginning. The reset of that song that was cut, reset, cut, reset ... it was on purpose. I know it was repetitive, but there was the idea we always had to reset the phone, and I just wanted to acknowledge that.

(gesturing towards the triplet sculpture scene) ... this looked like it was all distinctive bodies, but ten seconds ago she was morphing out of it. There's also something interesting in that which we didn't have time to choreograph for. I think sometimes the morphing creates this horror or this alien body.

(talking on Cloe giving Immie specific direction) ... so here for Imogen ... because of the orientation of the cameras, we had very different directions. In a way, our north was not the same ... I was so focusing on [the hologram] that I forgot to look at what she was actually doing. So I was guiding her through what I was seeing, but she didn't know what I was seeing. So I guess there's a lot to learn about how you communicate with the dancer, and obviously I failed.

M: How could you have made Immie's job easier?

C: Well, again it's a work of imagination, and I should have imagined myself, projected myself into her direction. You know sometimes I work with a numbering of the space ... so I should have told her about the numbering instead of telling her ″right″, ″left″, I should have gone ″towards three″, or ″towards four″ ...

... maybe there's something about filming certain sections and potentially projecting them in a 2d as part of the bigger work ... maybe it's a backstage moment or maybe there's a screen with just one seat. Where a person sits down and just watches that after having the experience with the glasses ...

(talking about the triplet scultpure) You know I say it depends on where people will be, but we actually managed to have this scene pretty tight every single time. So that's a win. From wherever you were standing it was interesting ... I also like with the projection, it wasn't in the reflection of the mirror here ...

... Something for the future to think about is also the costumes. Because see here the black is not see-through at all. It's a different type of black, and I quite like the red because I mean I always put a bit of red in every show ... but because it's not too red with the color treatment in the in the glasses ... it's important because there was some some colors we tried which actually didn't work at all because they disappear. They are too close to black, even a green that looks nice real ... also seeing the skin is very nice because you get the definition and you need the definition ... So yeah there's something about the the fabric of the clothing and also the colors to consider.

M: Could you elaborate a little bit more on how the quality of the cameras influenced your decisions?

C: I think maybe when the glasses get developed further I might not use the red pants because the red might just pop too much ... With the Netherlands we were in there with those Adidas pants ... suddenly you just see the brand, and it's so intimate with the glasses you don't want that to happen here.

I think we really need to consider what we wear. Especially the material that it's made of because your black t-shirt disappearing... I think David was wearing blue pants and it tended to disappear as well which is why he removed his t-shirt and then we realized our skin is nice. You can see the muscle definition, so even more muscle definition is nice because that makes the person ... or non muscle definition depending on the body ... but just to see the body is nice.

I think the glasses were making everything lighter. Blander, maybe ... like Imogen's hair here looks very light. So I think the color grading [is] a thing to think about ... It could be the glasses is ... can you change it within the software ... you still want it to look natural unless you're trying to go for something unnatural.

Like when we were working towards the camera and it was creating those weird legs ... "the beast".

C: ... the fact that we kept the projector at the back, even though we didn't use it for the showing ... for me was important to keep it because we used it to understand the technology briefly... For me it was about not hiding what we had used to develop the work ... it [was] a definite creative decision. It's not like we do this because we can't deal with it any other way ... I think the audience then really have an insight into a creative process... maybe because we know it's developed from a research but I don't want to shy away from that because I think technology is a constant research, choreography is a constant research and why not actually embrace that.

... and you know you are also a performer within this. You're not just someone from the shadows. I think with technology we tend to just show the results. I like seeing your table on the side. . You are a performer in it ... and all this technology are other performers. It becomes the partners we play with.

When we do harness work, you see the harness ... so here you see the technology, the brain behind it.

Maybe this is not for every project, but with this it brings a bit of fragility to it ... it's a trial and error ... we create less expectation as well. It's like it doesn't have to be perfect because it actually can't be at this stage, and that's okay. That's what we're playing with.

M: Do you think that makes it less magical?

C: To me it makes it more magical. Maybe because I'm not technology person. And also we haven't really revealed the back space. It's just when people were coming in they could see all that, but the dream would be to reveal that back space or having them ... go wherever.

I went to see a show for Sydney fest ... it was about Antarctica ... and then at the end you could look at — I mean some people were very nerdy I guess and were really interested in the technology of this ... I wasn't particularly ... though I still looked — it was like "oh wow all these little circuits". That was the magic because I didn't understand it, but I really enjoyed looking at it. But other people were really looking at it because they were trying to understand. So if anything it just makes it more real.

C: I feel like also because with the digital for me — again, subjective — there is a lot of humanity within it. I find it becomes really sanitised and perfect ... Part of me wanted to fight against the digital a little bit and find the humanity within it. So seeing the cables, seeing the projection that we always tried to hide, I think for me was good to show it. It's like showing a body part, showing, you know, a dancer that can't do a move perfectly because it's just human. Well, the digital is just a machine. And in a way, the human body is a machine as well. It can break, it can fail. And that's okay. ... if we were to continue that and have like a composer as well or musician or whatever, I think it would need to be seen with all whatever technology, instruments he's using ... all the cables.

C:(ref:erencing the triplets) So yeah, here ... we'd have breath ... to make it more human instead of statuesque ...The statuesque works when the dancer's body is actually not moving as well. So that's where you can create the illusion ...

C:(ref:erencing changing the viewing angle of triplets) ... So, here is when I manipulated the audience and I really ask them "Can you please come to this diagonal?". I'm on it because sometimes I see things and I'm like "Am I the only one seeing this?" ... Maybe it's just interesting for me.

But that idea of as she comes towards the glasses, the real body (because of the filter or the glasses) ... she becomes quite hologram like. And that was nice for me because it was the ending as well. So she disappears and it's like the fragility of human existence in a way, or like the body is just an envelope, whatever you want to see in it. And a few people went there. How can you work in a way that this real body would treat that reality and it looks like it's a hologram? How can we do that? ... thats a good challenge for you ... it would be the digital taking over. A switch of power.

C: I did tend to stay away from the hologram a lot, I realize now ... I don't know why. Because it was creating such beautiful modern art from clothes....

M: Did it have anything to do with the limited field of view?

C: Maybe. And I think also because it was almost going further than the parameters that we had, and I think it was a next step type of thing. But realizing that through ... the combination of choreography, dance, technology, creating these sculptures, these ephemeral sculptures ... or long-term, as long as you have the glasses, you know. Like a fake, ephemeral ... or exotic moment, which then switches your role as a choreographer. I felt like a bit like a sculptor as well, actually ... it's a switch of power within my own role. Maybe for you as well, when you were changing the texture, you become a sculptor as well. So it ... becomes very interdisciplinary, in a way, thanks to the technology. The technology ... gives you another job in a way ... which is interesting because that wouldn't happen in the real world so much.

C: So, going closer, you could see the piece. But it was so abstract and I felt for now, we were not ready for such abstraction ... Because then I was taking the whole narrative or images that we had from the beginning into something else. So it was nice for people at the end because at the end they come close and experience the closeness and what it can do as well ... It was good that we didn't go further than that. But there is another step ... I think that was extremely difficult because you just you disappear [into abstract] very quickly.

C: ... So then what I'd say is that we go back to this manipulation of the audience ... How do we make them want to come closer without having to actually tell them. You can go close.

C:(ref:erencing the triplets frozen)... It's like a drawing as well. Because, through all this digital you know, there's a lot of things that we experience like oil and paper, in a way. So I think it just combines so many disciplines in one. It's quite extraordinary, what it does with just one pair of glasses and just ... four people working on it. ... it combines architecture, it combines the design. We did everything without those people ... so it's a huge potential for performance.

M: The technology failed quite a lot. Like, things floated around and shifted around and. But we kind of made it work as it was. ...

C: It would be better in the future if the tracking was solid, but I think some of these things that didn't work opened up possibilities or made us think.

C: ... the digital always want it to be perfect, but it's also something I'm interested in is the imperfection. It's the same in dance. I like the imperfection. Or when people make a mistake, I think that's when it's interesting. Because I think the mistake or the imperfection is what creates the provocation.

M: ... that is what I feel I would like to do as a technologist ... we have this imperfection which has now inspired some idea. What we really want is to get rid of the imperfection but have control over it entirely now. So say, you know, the pixelation at the back would be awesome if we could just control it. So they're never pixelated ever, unless we want them to be. But now that we understood that that's something that we want, we can improve the quality and then just take back control because we had limited control, which is what resulted...

M: The idea of using new technologies, like a lot of the frustrations that come out of it through lack of control. Yeah, but this it's also been beneficial. Like you say, human beings. Yeah.

C: Lack of control. But yeah, and as opposed to the first residency, I guess where we use all those shapes as well which. Sure. Yeah. I really like them too. And we went in different tangent, but I think maybe we could have a cue here and there added to it. But in a very sporadic minimalist because I think it's still fit as well. So there's the cube thing was very nice, but yeah, in a way also because we discussed about the interaction with David. And then I think it's a different way of thinking of interaction as well. Maybe for me, because I think David likes people to really interact deeply and I feel maybe with this it's more interacting in a subdued way because ... the interaction is over you through the glasses.

... we already afford them a lot more control just by putting the headset on and they can walk around.

B. Group Discussion

In the following excerpt, 'Q' denotes the question-asker, whether an audience member or the on-stage MC.

Q: Could you tell us a bit how this project started?

Director 1: Um, it started during COVID. Uh, and it was really interesting. At the beginning we were all in lockdown. We couldn't leave our homes ... or maybe just for exercise. And I thought, why don't I make a work with people who were within walking distance of my home. And Cloé lives nearby, and we thought to make a piece under a tree in a park, which is at the end of the show now, of course. And Matt lived within walking distance. Boris, he lives within walking distance now. And then after that — [after] this idea of something very close [we] thought why don't we expand out? And Katina was in lockdown in Queensland and I thought to work with her. And then of course we went "oh, we can work in a distance, so why don't we work a very long way away?" And then then the idea came to partner with the Netherlands. That's the short version.

Q: And then maybe a question for Katina and Cloé. Could you tell us a bit about what it is like to perform 16,000km away?

Performer 1: Uh, it's. It feels further away when there's a delay with the Internet. But it's been.

It's been quite special, actually, to to perform with someone on the other side of the world. I think especially during these times that's a pretty special thing. When, um. Yeah. We can't travel internationally and yeah, I found that pretty cool.

Performer 2: So it's interesting because there's another version Olympia experienced different from what we experienced tonight.

Everything made sense once Olympia was performing with us.

[Before that], we made choreography, we were exchanging a lot. I was sending videos to her, she was sending me videos. We were making the work that way... and when we had our 'runs' [rehearsals together], that's when we understood the relationship that we needed to create between her virtually on the wall and then us...

And then tonight, we were the ones. I mean, we don't have lights... we have this video... and it's a very different awareness because we have to look at the monitors, we have to be exactly in the same spot every night...

...it's super interesting. It's a brain exercise and understanding the relationship. Um, yeah, I felt like tonight I worked a lot more with my brain... not "thinking"

Performer 1: That's true, because in the first version [with the audience in Sydney] we were looking at Olympia on the [projection]. But in this version [with the audience in Gronengen] we're ... imagining her in this space. It's kind of imaginary, but then when we see the monitor, we can see the interaction and know that that's there.

Q: (translated) So they're asking whether you have a public...

Director 2: ... they don't have public because it's very late at night. That's of course the problem with working with Australians. There's ten hours difference.

Q: Technically it would be possible?

Director 2: Well, we did it the other way around two weeks ago. They had the live performance in Brand X — it's a grand theatre in Sydney. And then we were in an office at 8 o'clock in the morning.

Q: So maybe you could tell more about that there are two different performances?

Director 2: I think we made two completely different live versions.

Director 1: Yeah, it's interesting to make ... within two weeks ... two very different versions of the same show, and partly that was because we were in a theatre that couldn't set up anything in. So that created one problem. But there's this feeling that with the digital work, you can change it quite rapidly and make quite different versions quite fast. And so it actually quite a creative platform to work where you can make many iterations in a very short period of time with just a few building blocks. I'm really interested to see "what next?", where we take it next and how far we can take this "fluid" work. I guess you would say it's very fluid.

Director 2: I had a panic attack the other day when you told me you wanted to add a third company in Mexico.

Director 1: And then after that I was thinking the Moon maybe...

Q: I have a question really for the the dancers ... I don't know very much about dance, but I'm trying to imagine for the dancers ... it begins quite prominently, very specific. It's marked space, it's very precise. And the distance that you're being put through whith various clever bits of software ... it begins very technically ... Is there within this process a way that it becomes intuitive between you in another way?

Performer 2: When we were on the table, that part for me is very much intuitive. I didn't have to look at the monitors at all, because it hadn't changed much from the previous version we'd done. I could figure out where Olympia was, so then it was just about adjusting my gaze to where I believed he was. I mean she didn't move too much, it was more the angle of her body, or when to turn around, or when to connect with her. So this was very intuitive. I guess the beginning, because we follow each other, that was the hardest part ... or ... it's kind of intuitive as well, I guess.

Performer 1: It took some time at the start to find the timing with each other. But I think after a while we started to feel what really worked. So that was really nice ... finding that timing together when we wanted to be in time and when the choices were to be in a cannon or something.

Performer 2: Because there's the music. So once you understand what's happening with the graphics ... and the music ... it feels like in the room even though she's not. It's really weird but it feels like she's present.

Director 1: One thing I will say is when we do the show or we do a rehearsal without Olympia, it feels very flat. It feels very, very flat. But as soon as ... the three dancer's are together, there's a magic that happens. And that's really interesting because they're 12,000km apart or whatever it is ... but you feel the presence of the third person.

Performer 3: I think it was very different because they are two dancers , so they made material together, whereas I was by myself. And because we didn't make movements together, it was not a trio in that sense where we made material together it was actually very different material [we each brought].

I think the intuitive part was the energy, because actually in the box I don't "dance" much. It's more angles of the body ... How you send focus into this blank space that then gets transferred to them. That's I think, the intuitive part. So yeah, it's, it's more listening than anything because I can do a movement and it looks nice on the screen, but it can really easily be disconnected ... I'm also looking at the monitor and I'm looking at that because I'm also there ... so it's just more energy than anything else ...

Director 2: Let's make a transition to Boris and Matt because they're also part of the choreography. Maybe you should say something. Say something ... nerds don't talk.

Tech 1: Is there a question?

Director 2: Just say something! What did you do in this project?

Tech 1: Oh, very good. What did I do? Well Matt and I were working in our respective software packages attempting to get the point-clouds distributed throughout this network of cameras and control systems to get something that's usable for the project essentially as well as develop aesthetic looks that were, you know.

Tech 2: Yeah I think a lot of the main challenges were with getting the cameras to pick up something that's usable and able to be put in this virtual space that can be shared. And then making these virtual environments that surround the dancers.

Tech 1: I think it's great working with with the WERC team as well.

Q: How do you work together? How does that work?

Tech 1: Frantically ... attempting to ensure that things don't crash. There's a lot of troubleshooting that's going on most of the time in terms of getting all this information around and distributed and connected between us just within our own computer systems and also then between our computer systems and their computer systems, while also trying to be creative at the same time.

Q: Yes exactly — because we're talking now much about the technical side, but there must be also the artistic side that you exchange?

Tech 1: Because of our really limited time that we had that we got to spend to connect together, a lot of the time was spent troubleshooting technical stuff, to be honest, because there was just so much to overcome.

So I think when we're talking about intuitive collaboration, I think it was mainly to do with us being intuitive off of what they were presenting to us and reacting to that and hopefully — I think probably — vice versa. So there wasn't really a lot of, a lot of direct creative time that we got to spend. I think we just sort of went and ran with with ideas that that people offered up. You know, the work offered up and and we kind of came together.

Tech 2: And I think ... working around the time zones. We would spend eight hours working and developing ... we would come up with a whole set of fresh stuff and then go to sleep and then come back the next day and have a response to that. And we kind of went through 2 or 3 weeks development of just doing that back and forth daily.

Tech 3: I think what's good to mention too is while the dancers communicate visually (they see the video streams), we have multiple chat windows open where we talk about cues, what's going wrong, which server went down, where to pick it up. So that's what we're doing constantly there on a chat window. And at the same time we have two technical setups that are well, not identical but that are comparable. So the same stuff is running on the two sides of the world and then they are being exchanged. So all of the manipulations you see of the point-clouds dissolving or changing ... on the gauze and on the screen, that's something Joachim and Jelle are doing there, but that's something they are also doing at the same time in Australia. So we have a chat with technical troubleshooting stuff, like "how stuff works" and "do they hear the sound that we just sent them" and stuff like that. And at the same time they are seeing what we're doing and trying to have a visual response to this and trying to match up the visual worlds together.

Q: Me as audience. I didn't get that.

Tech 3: You're supposed to not get that, right!

Q: It's really like you are the ones who control the whole video department. We don't see video on their side. So it's you to control. And what is the purpose then that they also have control?

Tech 3: What they need to do the local manipulations of the two dancers there because we cannot send ... it's too much information to send over the Internet, so they're doing some of the manipulations there already, and then sending it ... so they're like pre-setting stuff so we can incorporate it. And then what we do is we decide what we put to the projectors. But actually the weird thing is, and this I totally agree with you, there is a big world you don't see, which is what happening in in the middle of these two places where the the streams collide and where two computers on the other side of the world are calculating similar things so they can transmit together.

Tech 3: The first show was us *giving* to Australia. This one is Australia giving to us. And the next step would be to have the real one.

Q: Yeah, that would be better.

Q2: I think we should wrap up because we have to send you guys to sleep. I mean, it's past midnight.

Director 1: Yeah, we have to do our bump-out now .. we have to break our set down. So yes, late night for us. It's a pleasure working, a real pleasure working with you all. Really, really grateful.

Tech 3: Same here.

C. Publications

During the course of this PhD, I co-authored the following publications. These may discuss works or concepts in this thesis and some of their text may appear verbatim in this work, despite not being explicitly referenced.

Hughes, M. (2020). URack: Audio-visual composition and performance using Unity and VCV Rack. *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Hughes, M., Garcia, J., Wilcox, F., Sazdov, R., Johnston, A., & Bluff, A. (2020). Immerse: Game engines for audio-visual art in the future of ubiquitous mixed reality. *Proceedings of the 5th International Conference on Live Interfaces*.

Ilsar, A., & Hughes, M. (2020). Harmony across music, visuals and movement in a new audio-visual gestural performance. *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pp. 625–630.

Ilsar, A., & Hughes, M. (2021). AirStream: A collaborative gestural virtual performance. *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Ilsar, A., Hughes, M., & Johnston, A. (2020). NIME or mime: A sound-first approach to developing an audio-visual gestural instrument. *Proceedings of the International Conference on New Interfaces for Musical Expression*.

References

Alais, D., Newell, F., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and perceiving*, *23*(1), 3–38.

Alves, B. (2005). Digital Harmony of Sound and Light. *Computer Music Journal*, *29*(4), 45–54.

Atkinson, P. (2007). *Ethnography: Principles in practice*. Routledge.

Attride-Stirling, J. (2001). Thematic networks: An analytic tool for qualitative research. *Qualitative research*, *1*(3), 385–405.

Auslander, P. (2006). The performativity of performance documentation. *PAJ: A Journal of Performance and Art*, *28*(3), 1–10.

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments*, *6*(4), 355–385.

Basu, A., Saupe, C., Refour, E., Raij, A., & Johnsen, K. (2012). Immersive 3dui on one dollar a day. *2012 IEEE symposium on 3D user interfaces (3DUI)*, 97–100.

Beaudouin-Lafon, M. (2004). Designing interaction, not interfaces. *Proceedings of the working conference on Advanced visual interfaces*, 15–22.

Berthaut, F., Martinez, D., & Hachet, M. (2015). Reflets: Combining and Revealing Spaces for Musical Performances. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 116–120. https://hal.inria.fr/INRIA-BORDEAUX/hal-01136857v1

Bluff, A., & Johnston, A. (2017). Creature:Interactions: A Social Mixed-Reality Playspace. *50*(4), 360–367. https://doi.org/10.1162/LEON

Bluff, A. J. (2017). *Interactive art, immersive technology and live performance* (Doctoral dissertation). University of Technology, Sydney. http://hdl.handle.net/10453/120340
viewed 20 February 2018

Boehler, W., Vicent, M. B., Marbs, A., et al. (2003). Investigating laser scanner accuracy. *The international archives of photogrammetry, remote sensing and spatial information sciences*, *34*(Part 5), 696–701.

Boehm, B., & Turner, R. (2009). *Balancing Agility and Discipline: A Guide for the Perplexed* (7th). Addison-Wesley Professional.

Borgdorff, H. (2012). *The conflict of the faculties: Perspectives on artistic research and academia*. Leiden University Press. http://www.oapen.org/search?identifier=595042

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77–101.

Brechner, E. (2015). *Agile project management with kanban*. Pearson Education.

Breslauer, J. (1992). Theater on a chip: George coates' 'invisible site' mingles performers and 'live' computer images in a symbiotic work that redefines the state of the art. *Los Angeles Times*. https://www.latimes.com/archives/la-xpm-1992-01-12-ca-466-story.html

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.

Bui, M., Chang, L.-C., Liu, H., Zhao, Q., & Chen, G. (2021). Comparative study of 3d point cloud compression methods. *2021 IEEE International Conference on Big Data (Big Data)*, 5859–5861.

Candy, L. (2006). *Practice based research: A guide* (tech. rep.). Creativity & Cognition Studios, University of Technology. Sydney. http://research.it.uts.edu.au/creative/eae/intart/pdfs/PBR%20Guide-1.1-2006.pdf

Candy, L. (2011). Research and creative practice. In L. Candy, E. Edmonds & R. Ascott (Eds.), *Interacting: Art, research and the creative practitioner* (pp. 33–59). Libri Publishing Ltd Farringdon, UK.

Coutrix, C., & Nigay, L. (2006). Mixed reality: A model of mixed interaction. *Proceedings of the working conference on Advanced visual interfaces*, 43–50.

DeSantis, L., & Ugarriza, D. N. (2000). The concept of theme as used in qualitative nursing research. *Western journal of nursing research*, *22*(3), 351–372.

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, *9*(6), 719–721.

Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence [Copyright - Copyright Human Factors and Ergonomics Society Sep 1998; Last updated - 2023-03-01; CODEN - HUFAA6]. *Human factors*, *40*(3), 354. http://ezproxy.lib.uts.edu.au/login?url=https://www.proquest.com/scholarly-journals/telepresence/docview/216450988/se-2

England, J. (2018). Science of Pepper's ghost illusion. *Cosmos Magazine*. https://web.archive.org/web/20230503221550/cosmosmagazine.com/science/physics/the-science-behind-the-peppers-ghost-illusion

Farivar, C. (2012). Tupac "hologram" merely a pretty cool optical illusion. *Ars Technica*. https://www.theverge.com/2017/5/2/15477520/sixense-stem-oculus-rift-vr-controller-kickstarter-problems

Florwick, J. (2022, July 14). *Cisco 6 ghz planning and mobility innovations update* [Presentation]. Mobility Field Day 7, Silicon Valley. https://youtu.be/VfFdesxz29w?t=579

Fofi, D., Sliwa, T., & Voisin, Y. (2004). A comparative survey on invisible structured light. *Machine vision applications in industrial inspection XII*, *5303*, 90–98.

Fox, R. (2007). Sound and light signal simultaneity. *Filter Magazine*, *66*. http://filter.org.au/issue-66/sound-and-light-signal-simultaneity

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., & Marín-Jiménez, M. J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, *47*(6), 2280–2292.

Gaver, W. (2012). What should we expect from research through design? *Proceedings of the SIGCHI conference on human factors in computing systems*, 937–946.

Godard, C., Mac Aodha, O., Firman, M., & Brostow, G. J. (2019). Digging into self-supervised monocular depth estimation. *Proceedings of the IEEE/CVF international conference on computer vision*, 3828–3838.

Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T., & Tabatabai, A. (2020). An overview of ongoing point cloud compression standardization activities: Video-based (v-pcc) and geometry-based (g-pcc). *APSIPA Transactions on Signal and Information Processing*, *9*, e13.

Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., Koller-Meier, E., Svoboda, T., Van Gool, L., Lang, S., Strehlke, K., Moere, A. V., & Staadt, O. (2003). Blue-c: A spatially immersive display and 3d video portal for telepresence. *ACM Trans. Graph.*, *22*(3), 819–827. https://doi.org/10.1145/882262.882350

Guo, K., Lincoln, P., Davidson, P., Busch, J., Yu, X., Whalen, M., Harvey, G., Orts-Escolano, S., Pandey, R., Dourgarian, J., et al. (2019). The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (ToG)*, *38*(6), 1–19.

Gupta, M., Yin, Q., & Nayar, S. K. (2013). Structured light in sunlight. *Proceedings of the IEEE International Conference on Computer Vision*, 545–552.

Hamilton, R. K. (2011). UDKOSC: An Immersive Musical Environment. *Proceedings of the International Computer Music Conference*.

Han, B., Liu, Y., & Qian, F. (2020). Vivo: Visibility-aware mobile volumetric video streaming. *Proceedings of the 26th annual international conference on mobile computing and networking*, 1–13.

Hanington, B., & Martin, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Publishers.

Hintjens, P. (2013). *Zeromq: Messaging for many applications*. " O'Reilly Media, Inc."

Huang, X., Mei, G., Zhang, J., & Abbas, R. (2021). A comprehensive survey on point cloud registration. *arXiv preprint arXiv:2103.02690*.

Ikeshiro, R. (2012). Audiovisual harmony: The realtime audiovisualisation1 of a single data source in construction in zhuangzi. *Organised Sound*, *17*(2), 148–155.

Ilsar, A. (2018). *The AirSticks: A New Instrument for Live Electronic Percussion Within an Ensemble* (PhD Thesis). University of Technology Sydney. http://hdl.handle.net/10453/125647

Ilsar, A., Havryliv, M., & Johnston, A. (2013). The airsticks: A new interface for electronic percussionists. *Proceedings of the Sound and Music Computing Conference, SMC 11*, 220–226.

Ip, C. (2017). 'Flying Lotus in 3D' is a jam session between music and holograms. *Engadget*. http://web.archive.org/web/20221208043815/engadget.com/2017-11-25-flying-lotus-in-3d-concert-interview.html

Jacquet, Y., Peeters, J., & Vaquié, T. (2007). 3Destruct. http://blog.antivj.com/2012/3destruct-scopitone/%20http://blog.antivj.com/2014/3destruct-berlin-atonal-2013/

Johnston, A. (2013). Fluid Simulation as Full Body Audio-Visual Instrument. *NIME 2013: New Interfaces for Musical Expression*, 5. http://nime2013.kaist.ac.kr/program/papers/day1/paper2/151/151%7B%5C_%7DPaper.pdf

Johnston, A. J. (2009). *Interfaces for musical expression based on simulated physical models* (Doctoral dissertation). University of Technology, Sydney.

Kauff, P., & Schreer, O. (2002). An immersive 3d video-conferencing system using shared virtual team user environments. *Proceedings of the 4th international conference on Collaborative virtual environments*, 105–112.

Kepner, L. S. (1997). Dance and digital media: Troika Ranch and the art of technology. *Digital Creativity*, *8*(1), 11–19.

Kowalski, M., Naruniec, J., & Daniluk, M. (2015). Livescan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors. *2015 International Conference on 3D Vision*, 318–325.

Krueger, M. W., Gionfriddo, T., & Hinrichsen, K. (1985). Videoplace—an artificial reality. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 35–40.

Kurillo, G., & Bajcsy, R. (2013). 3d teleimmersion for collaboration and interaction of geographically distributed users [Copyright - Springer-Verlag London 2013; Document feature - ; Last updated - 2014-08-30]. *Virtual Reality*, *17*(1), 29–43.

Laga, H., Jospin, L. V., Boussaid, F., & Bennamoun, M. (2020). A survey on deep learning techniques for stereo-based depth estimation. *IEEE transactions on pattern analysis and machine intelligence*, *44*(4), 1738–1764.

Leonard, N. E., Young, G., Hochgraf, K., Swain, D., Trippe, A., Chen, W., & Marshall, S. (2012). In the dance studio: Analysis of human flocking. *2012 American Control Conference (ACC)*, 4333–4338.

Mao, J.-Y., Vredenburg, K., Smith, P. W., & Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*, *48*(3), 105–109.

Masoumian, A., Rashwan, H. A., Cristiano, J., Asif, M. S., & Puig, D. (2022). Monocular depth estimation using deep learning: A review. *Sensors*, *22*(14), 5353.

Mazzanti, D., Zappi, V., Caldwell, D., & Brogni, A. (2014). Augmented stage for participatory performances. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 29–34. http://www.nime.org/proceedings/2014/nime2014%7B%5C_%7D413.pdf

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

Mitchell, T., & Heap, I. (2011). SoundGrasp : A Gestural Interface for the Performance of Live Music. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 465–468. https://doi.org/10.5281/zenodo.1178111

Naugle, L. M. (2002). Distributed choreography: A video-conferencing environment. *PAJ: a Journal of Performance and Art*, *24*(2), 56–62.

Nguyen, N. T., McFadden, A., Tangen, D., & Beutel, D. (2013). Video-stimulated recall interviews in qualitative research. *Australian Association for Research in Education*.

Nijholt, A., & Traum, D. (2005). The virtuality continuum revisited. *CHI'05 Extended Abstracts on Human Factors in Computing Systems*, 2132–2133.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods*, *16*(1), 1609406917733847.

Oliveros, P., Weaver, S., Dresser, M., Pitcher, J., Braasch, J., & Chafe, C. (2009). Telematic music: Six perspectives. *Leonardo Music Journal*, *19*(1), 95–96.

Oughton, E. J., Lehr, W., Katsaros, K., Selinis, I., Bubley, D., & Kusuma, J. (2021). Revisiting wireless internet connectivity: 5g vs wi-fi 6. *Telecommunications Policy*, *45*(5), 102127.

Phillips, A. (2010). From Work to Practice. In M. Boon & G. Levine (Eds.), *Practice - whitechapel: Documents of contemporary art* (pp. 71–73). The MIT Press (Collection published 2018).

Robertson, A. (2017). The sixense stem was once the ultimate vr controller. where is it now? *The Verge*. https://www.theverge.com/2017/5/2/15477520/sixense-stem-oculus-rift-vr-controller-kickstarter-problems

Rowell, D. (2019). The spectacular, strange rise of music holograms. *The Washington Post Magazine*. https://web.archive.org/web/20220516023456/washingtonpost.com/magazine/2019/10/30/dead-musicians-are-taking-stage-again-hologram-form-is-this-kind-encore-we-really-want/

Shaw, J., & Botschuijver, T. (1975). Viewpoint. *9th Biennale de Paris, Musee d'Art Moderne, Paris, France*.

Shaw, J., & Botschuijver, T. (1981). Virtual sculpture. *Melkweg, Amsterdam, Netherlands*.

Shaw, J., Botschuijver, T., & Wellesley-Miller, S. (1970). Cloud (of daytime sky at night). *Sedelijk Museum, Amsterdam, Netherlands*.

Sheppard, R., Wu, W., Yang, Z., Nahrstedt, K., Wymore, L., Kurillo, G., Bajcsy, R., & Mezur, K. (2007). New digital options in geographically distributed dance collaborations with teeve: Tele-immersive environments for everybody. *Proceedings of the 15th ACM International Conference on Multimedia*, 1085–1086.

Silva, S. (2022). Gorillaz turn the world into a stage with augmented reality. *The Keyword Google Blog*. https://blog.google/products/google-ar-vr/gorillaz-maps-music-video

Skarbez, R., Smith, M., & Whitton, M. C. (2021). Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, *2*, 647997.

Spinuzzi, C. (2005). The methodology of participatory design. *Technical communication*, *52*(2), 163–174.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. MIT press.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1517.

Sutherland, I. E. (1968). A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, 757–764.

Tanaka, A. (2000). Musical performance practice on sensor-based instruments. *Trends in gestural control of music*, *13*(389-405), 284.

Tanaka, A., & Bongers, B. (2002). Global string: A musical instrument for hybrid space. *ICMC*.

Tölgyessy, M., Dekan, M., Chovanec, L., & Hubinskỳ, P. (2021). Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2. *Sensors*, *21*(2), 413.

van der Hooft, J., Vega, M. T., Wauters, T., Timmerer, C., Begen, A. C., De Turck, F., & Schatz, R. (2020). From capturing to rendering: Volumetric media delivery with six degrees of freedom. *IEEE Communications Magazine*, *58*(10), 49–55.

Vasudevan, R., Zhou, Z., Kurillo, G., Lobaton, E., Bajcsy, R., & Nahrstedt, K. (2010). Real-time stereo-vision system for 3d teleimmersive collaboration. *2010 IEEE International Conference on Multimedia and Expo*, 1208–1213.

Vovk, A., Wild, F., Guest, W., & Kuula, T. (2018). Simulator sickness in augmented reality training using the microsoft hololens. *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–9.

Walsh, L., Bluff, A., & Johnston, A. (2017). Water, image, gesture and sound: Composing and performing an interactive audiovisual work. *Digital Creativity*, *28*(3), 177–195. https://doi.org/10.1080/14626268.2017.1353524

Whitney, J. (1980). *Digital Harmony: On the Complementarity of Music and Visual Art*. Byte Books.

Williams, T., Szafir, D., & Chakraborti, T. (2019). The reality-virtuality interaction cube. *Proceedings of the 2nd international workshop on virtual, augmented, and mixed reality for HRI*.

Wobbrock, J., & Kientz, J. (2016). Research contributions in human-computer interaction. *Interactions*, *23*(3), 38–44. https://doi.org/10.1145/2907069

Wu, C.-H., Hsu, C.-F., Kuo, T.-C., Griwodz, C., Riegler, M., Morin, G., & Hsu, C.-H. (2020). Pcc arena: A benchmark platform for point cloud compression algorithms. *Proceedings of the 12th ACM International Workshop on Immersive Mixed and Virtual Environment Systems*, 1–6.

Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & psychophysics*, *67*(3), 531–544.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in hci. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 493–502.