



UNIVERSITY
OF SUSSEX

Doctoral Thesis DRAFT

Material, Embodied, and Spatial Relations in Augmented Reality

An Exploration of AR as a Medium for Musical Performance and Experience

*A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy in the*

Department of Music
School of Media, Arts and Humanities
University of Sussex

Author:

Sam Bilbow

Supervisors:

Dr. Chris Kiefer

Dr. Cécile Chevalier

Prof. Jamie Ward

28th October 2022

For Hanna, George, Diane, and Otto

ACKNOWLEDGEMENTS

This doctoral research was funded by the Leverhulme Trust Doctoral Scholarship Programme: "Sensation and Perception to Awareness". Directed by Prof. Anil Seth, and Prof. Jamie Ward.

I, Sam Bilbow, hereby declare that this thesis has not been and will not be, submitted in whole or in part to another university for the award of any other degree.

Brighton, UK,

28th October 2022

Sam Bilbow

University of Sussex
School of Media, Arts and Humanities
Department of Music

DOCTORAL THESIS

Material, Embodied, and Spatial Relations in Augmented Reality

by Sam Bilbow

ABSTRACT

This paper resituates multisensory augmented reality (MSAR) as an artistic medium for the creation of interactive and expressive works by computational artists. If an AR system can be thought of as one that combines real and virtual processes, is interactive in real-time, and is registered in three dimensions; why do we witness the majority of AR applications utilising primarily visual displays of information? In this paper, I propose a practice-led compositional approach for developing 'MSAR Experiences', arguing that, as an medium that combines real and virtual multisensory processes, it must explored with a multisensory approach. The paper further outlines the study methods that I will use to evaluate the developed experiences. The outcome of this project is the practice-led method as well as MSAR hardware, software and experiences that are developed and evaluated.

CONTENTS

1	Introduction	1
1.1	Multisensory AR as a Medium for Embodied Musical Performance	1
1.2	Personal Motivations	2
1.3	Key Research Questions	3
1.4	Aims & Objectives of the Thesis	4
1.5	Research Methods	4
1.6	Outline of Chapters	4
2	Literature Review	7
2.1	The Augmented Reality Interface	7
2.2	Augmented Reality in the Arts	19
3	Theoretical Framework	25
3.1	Knowledge and Aesthetic Experience	25
3.2	The 4E Approach to Experience	30
3.3	Augmented Materiality	34
3.4	Augmented Embodiment	46
3.5	Augmented Space	51
3.6	Conclusion	59
4	A DIY Approach to AR in the Arts	63
4.1	Research Methods	63
4.2	Resistance in Practice	65
4.3	Design Patterns for AR Art and Music	66
4.4	AR Experience Studies	68
5	The area~ System	69

5.1	Computational 'ARt' with area~	69
5.2	The area~ system	71
5.3	Study	75
5.4	Discussion	79
6	Evaluating polaris~	83
6.1	Design Framework	83
6.2	Study Design	87
6.3	Participant Feedback	89
6.4	Conclusion	95
6.5	Ethics Statement	97
7	Experimental A/V AR Performance @ The Rosehill	101
8	Discussion	103
9	Conclusion	105
A	Appendix: 5 The area~ System	107
B	Appendix: 6 Evaluating polaris~	109
C	Appendix: 7 Developing an AR Performance Practice	111
	Bibliography	113

INTRODUCTION

1

*I dream of instruments obedient to
my thought and which with their
contribution of a whole new world of
unsuspected sounds, will lend
themselves to the exigencies of my
inner rhythm*

(Varese and Wen-chung, 1966)

1.1 Multisensory AR as a Medium for Embodied Musical Performance

In the last twenty years, computational technology has become more and more expressive, the systems we engage with becoming increasingly interactive. Due to this, the arts (especially forms of sound-driven digital art) have embraced new technologies in tandem with, and often contributing to their development. This has lead to more human-centred methods of designing tools of digital art creation, moving away from black-boxed (undisclosed to the audience) interactions, towards more performative and suggestive interactions, which often include the audience. Among some of the technologies that have seen nascent in the arts is XR (the X stands for nothing and everything, the R stands for Reality), a group of technologies that promise to radically rethink the way we approach reality.

The category of XR comprises the technologies of: Virtual Reality (VR) - technology that promises to immerse us in a completely synthetic virtual world; and Augmented Reality (AR) - technology that promises to immerse us in our own real world through the computational mediation of virtual processes. This leads to the augmentation, diminishment, hybridisation, or extension of our own experience of reality, blurring and perhaps dissolving the line between our virtual and real selves. The present thesis makes no distinction between 'Mixed Reality' (MR)

and Augmented Reality, and adopts AR as an umbrella term for all contemporary (and mostly marketing) uses of the term MR.

AR, as mentioned, has seen sporadic use in the arts in the last twenty years. Its use in other disciplines, such as medical science, manufacturing and repair, annotation and visualisation, and the military (Azuma, 1997), have all been typically characterised by its use as technology that aids in the **visual overlay and alignment of virtual graphics** (text, image, animation, video) onto our **real world environment**. Within these fields, this has a potential myriad important uses, from training junior doctors in dummy procedures by overlaying the steps of a difficult surgery, to allowing three-dimensional networked collaboration across vast distances. Within the arts, its use has also been typically limited to this visual overlay approach. Despite being only a fraction of the possible forms of AR, this view of AR dominates the market of consumer AR devices, from head-mounted, to handheld and projective technologies, the majority deal with overlaying visual information. By taking a do-it-yourself (DIY) approach to designing and implementing custom digital musical tools through hardware and software, the present practice-based research thesis approaches AR through a multisensory lens to designing rich, expressive experience. The importance of such an approach is in its ability help redefine, and understand more about the technologies with which we are choosing to mediate our daily lives with, and also about how our own perception of digital media is affected through this embodied usage.

1.2 Personal Motivations

My own motivations for taking this multisensory, embodied approach to designing for AR is in the potential expressivity of the tool in creative rich, sensory artwork, which seeks to teach us more about the way we perceive ourselves, others, and our environment. Coming from a background in music technology, specifically in developing and evaluating digital music tools and instruments, the nature of typical ‘visual overlay’ AR is incompatible with its use as a digital music tool, and as such, it seems not only possible, but reasonable for it to be redefined to encapsulate all of the senses. In wanting to foster embodied experience between audience members (collaborative expression), AR offers a unique way of mediating and co-constructing (with them) the space in which they embody being and knowing. Thus, my motivation is also within the furtherment of our understanding this hybrid space, in which unique relationships between virtual and real processes are played out.

1.3 Key Research Questions

AR has seen nascent implementation as a creative medium in the arts, and more specifically music composition and performance. Therefore, the research areas and questions for the present thesis are based in the intersection between the arts, technology, sensory perception:

1. What is AR's genealogy, historical context, its contemporary forms, trends, exceptions, and definitions?
 - a) When did AR begin as a technology?
 - b) From which disciplines did AR begin?
 - c) What forms does AR take today?
 - d) Why is AR interesting and important today?
 - e) What definition of AR is appropriate for the arts?
2. What theoretical underpinnings can we use to explicate AR experiences within an artistic context?
3. What are AR's boundaries and affordances as a medium, are there problems it poses artists and musicians?
 - a) What does augmentation mean in AR, is this too restrictive a term?
 - b) What realities are typically augmented by AR?
 - c) What is the relationship between virtual and real content in AR?
 - d) Does AR tend towards an aesthetic?
4. What is the resultant experience for participants and audiences (or 'immersants') of an artistic AR experience?
 - a) What is reality for immersants of such experiences?
 - b) What does it mean for AR to modulate the sensory experience of an immersant?
 - c) Do AR immersants engage with physical and digital 'space' in novel ways?
5. What might a future corpus of artistic AR experiences and digital music instruments look, sound, and feel like?

1.4 Aims & Objectives of the Thesis

As well as answering and contributing understanding to the above research questions, this thesis aims to:

- Conduct a thorough literature and practice review of the field of AR within the arts
- Contribute to the understanding of the ‘tools of use’ that arise from ARs application within the arts
- Apply theories of space, materiality, and embodiment to contemporary AR design, use, and evaluation.
- Contribute to the practice field of AR arts by designing and evaluating multisensory AR tools for creative and collaborative expression.
- Contribute to the understanding of how such tools might affect our perceptions of self, others and our real, virtual, and hybrid environments.

1.5 Research Methods

The research methods used to answer my research questions, and achieve my aims and objectives above involve taking a practice-based approach to designing multisensory digital musical instruments using AR. Involved too, will be qualitatively (coded interview), and quantitatively (survey of experience) evaluating these instruments and tools, and iteratively redesigning them to provide rich and immersive experience.

1.6 Outline of Chapters

Chapter two starts with an in-depth history of AR as an emerging technology within the field of Human Computer Interaction (HCI), examining the historical barriers to usage such as portability, price that have led towards its popularity in a variety of forms today. Among these forms, the majority deal with simple and descriptive visual additions in the form of text, image, animation or video overlay. I make a case for ‘multisensory’ AR, i.e. the designing of content or experience for multiple senses to counter not only this bias toward visual design, but also as a method of attaining richer and more novel experiences of AR. This case is made by examining current advancements in multisensory HCI applications and enabling technologies. In the second part of the chapter, I posit through a review

of contemporary art practice, the usefulness of AR use within the arts in contributing to the designing and evaluation of these tools, enabling new aesthetic and multisensory experiences, collaborative expression, and agency.

In chapter three, I identify three lenses through which to view, not only the potential of modern-day applications of multisensory AR art, but also the potential of the medium itself. The first is through considering the intervention of AR processes in real and virtual spaces, how this effects the construction of these spaces, the implications this in-turn has for the type of art that is created, and the privacy and security of people co-located in these spaces. The second is through a consideration of the materiality and form of AR processes for both designer, user, and audience; not only by focusing on the tangibility and aesthetics of intervening virtual processes, but by evaluating the current hierarchies and structures that trickle down into mass-market AR technologies. The last is through application of current theories of radical enactivism / phenomenology in order to examine the experience of embodiment in AR processes.

Chapter four contains the methodology through which I address my research questions, this is separated into practice, study, data collection, and analysis methods. The first, practice methods, outlines the design principles for my approach to developing rich multisensory AR experiences (MSAR Experiences) through MSAR Instruments and MSAR Environments. These experiences vary in sensory intensity, engagement methods, and content. Snippets describe small-scale clip-like experiences, whereas Scenes describe medium-scale experiences, and Spaces in-turn describe experiences that take up the space of a room and involve an audience rather than a sole user. The second, reviews the study methods these AR experiences pass through in chapters five to seven; ranging from autobiographical design, to individual studies implementing iterative design, to a group audience study in the form of a large-scale AR installation. The last two sections reviews the data collection and analysis methods applied across the three studies.

In chapter five, I outline the development and evaluation of the ‘area’ system. ‘area’ enables users to record, manipulate, and spatialise virtual audio samples or nodes around their immediate environment. Through a combination of ambisonics audio rendering and hand gesture tracking, this system calls attention to the ability of non-visual AR, here, audio AR, to provide new aesthetic experiences of real and virtual environments. Through an autobiographical design study, this pilot study proposes that rich experience can result from non-visual AR systems.

Chapter six [study under construction]

Chapter seven [study under construction]

Chapter eight briefly reviews my theoretical framework and research methods before demonstrating the take home findings of my studies in order, and relative to my research questions. It also contextualises findings in the bigger picture of each research field. It details the contributions of the thesis, and offer concrete examples of how knowledge and practice has changed due to these contributions. It then provides future research areas related to the studies, and details any limitations of the methods and studies.

LITERATURE REVIEW

2

(Bilbow, 2022)

This literature review takes a retrospective view at the history of augmented reality technologies. In laying clear the origination of these technologies within predominantly visual fields of research, it allows for a critique of preconceived notions about the definitions of what AR is, and the boundaries of how it is used today. The review outlines the contemporary forms, sensory display techniques, and methods taken to mediate our reality with virtual processes. In sketching out this wide variety of possible arising interactions, I make clear the benefit of considering AR as a medium for expressive computational artwork, namely in its inherent ability to approach technology from a creative and inclusive perspective (not just from the perspective of viewing AR as a visual overlay tool). I then touch on examples of AR art that make much more creative use of the materiality of AR in this way.

2.1 The Augmented Reality Interface

2.1.1 History of AR

The earliest conceptions of what we would today call AR began their development in 1965 with Ivan Sutherland's 'Ultimate Display' (1968). Sutherland's project involved hypothesising and developing a computer interface that could serve 'as many senses as possible'. He speculated that this interface could 'serve as a looking-glass into the mathematical wonderland constructed in computer memory'. By 1968, Sutherland had constructed what he called a 'head-mounted three dimensional display', which was affixed to the ceiling of his lab with mechanical tracking arm that measured the position and orientation of the display in the room, called 'The Sword of Damocles'. The head-mounted display provided

per-eye visual displays via mini-CRT monitors, whose content was updated in real-time, dependent on the tracked pose of the head. This resulted in the three-dimensional (visual) perception of a hybridised real and virtual space, see [Figure 2.1](#).



Figure 2.1: Sutherland's 'Ultimate Display'

Over the course of the proceeding thirty years, most display (visual, audio and haptic) technologies gained higher resolutions, whilst becoming ever more miniaturised, propelled by the general movement towards ubiquitous and wearable computing. In 1992, augmented reality (AR) was officially defined by Boeing engineers Thomas Caudell and David Mizell, as a technology that could be used to 'augment the visual field of the user with information' (1992). Their prototype device featured display and tracking technologies built on the work of Sutherland. The purpose of the Caudell and Mizell's device was to increase manufacture workers' efficiency through overlay of graphical wireframe instructions. Their device operated by tracking elements of the users environment and body, processing these in real-time, and then overlaying these instructions in the appropriate places dependent on the manufacturing process through a HUDset. The HUDset, commonly referred to today as a head-mounted display (HMD) operated by updating the displays output in real-time dependent on the position and orientation (pose) of the participant. This had the effect of the graphical overlay seeming 'fixed', or what is referred to now as 'registered' or 'aligned', on top of a specific point in the real world. For Caudell and Mizell, this display and tracking solution is what leads to the 'technology' of AR being possible, and for them, it is what enables the context-aware instructional overlay of their application.

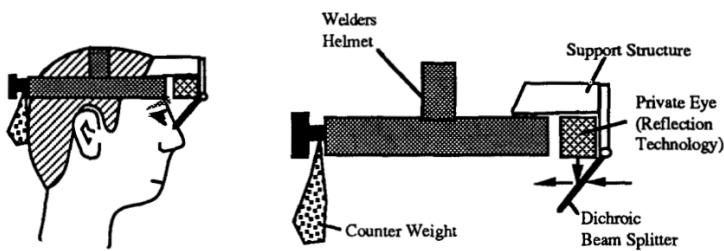


Figure 2.2: Caudell and Mizell's device

Shortly after this original definition, Milgram and Kishino conceptualised the 'Reality-Virtuality Continuum' (1994), which aimed to consolidate similar efforts across disciplines that were aiming to augment or virtualise 'reality'. The continuum (Figure 2.3) defines two outer bounds of 'Real Environments' on the left (which exist as physical and tangible matter), and 'Virtual Environments' on the right (that exist solely as digital bits), and everything within was to be classed as 'Mixed Reality' (MR). Within MR, use-cases closer to the 'Real' were regarded as 'Augmented Reality' (AR), and use-cases closer to the 'Virtual', were regarded as 'Augmented Virtuality' (AV). In separating these MR use-cases from the tele-robotics field of Virtual Reality / Virtual Environments (VR / VE), they proposed a clear framework for classifying works. Since both AR and AV make use of augmenting 'by means of real objects' in today's use cases, (e.g. by tracking and parameterising hand gestures, or tracking and generating spatial maps of the real environment), the difference between AR and AV in this definition, problematically, becomes the 'primary world' in which objects are doing the augmenting. Indeed, Milgram and Kishino make note of the possibility of this in their paper:

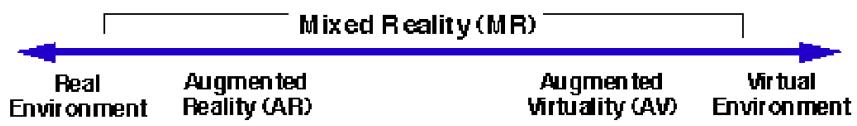


Figure 2.3: 'Reality - Virtuality Continuum'

'Of course, as technology progresses, it may eventually become less straightforward to perceive whether the primary world being experienced is in fact predominantly 'real' or predominantly 'virtual', which may ultimately weaken the case for use of both AR and AV terms'

They further specify that, despite this is the case, it should not affect the validity of the more general [Mixed Reality] term to cover the 'grey area' in the centre of the virtuality continuum, and also offer the term 'Hybrid Reality' for displays that blend AR and AV com-

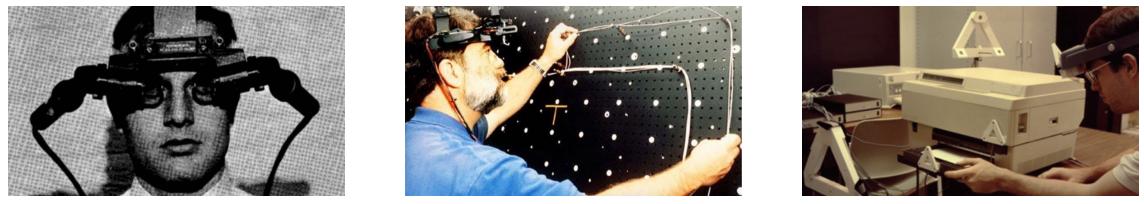
positively. Today, term ‘Mixed Reality’ is seen seldom apart from its use within Microsoft’s toolkit, and ‘Hybrid Reality’ even less so. Therefore, as previously mentioned, due to the rapid evolution of these technologies, the present thesis makes no distinction between Mixed, Hybrid, and Augmented Reality.

In 1997, Azuma proposed a specification for the definition of an AR system, and surveyed the five years of AR research since Caudell and Mizell’s original definition. In order to avoid limiting AR to specific technologies, they define AR as a ‘system’ that follows three characteristics:

- Combines real and virtual
- Interactive in real time
- Registered in 3-D

In this paper, he distinguishes between optical and video based approaches to AR and while both of these methods can and had been realised (when speaking strictly of visual display of information) via head mounted display technologies, Azuma makes further room in the definition for monitor based approaches too, allowing the definition to be broad enough to encompass a variety of AR use cases, methods and processes. Not soon after, Azuma et al., classify three categories of ‘display’: head worn, handheld and projective. For this reason, as well as the specification of three precise characteristics, Azuma’s definition of AR has become widespread.

These early definitions and conceptions of AR were typically built upon with applications involving display and tracking technology that resulted in the **visual overlay and alignment of virtual graphics onto our real world environment**. Whilst this view of AR is followed by the overwhelming majority of historical and contemporary uses of AR, it only makes up a small subset of the potential myriad interactions afforded by the technologies that enable AR. Sutherland, who arguably catalysed early practical research in this field, viewed this AR as a ‘window’ interface, that could serve ‘as many senses as possible’. Rosenberg, another early proponent of AR, developed the concept of ‘Virtual Fixtures’ (overlaid sensory information) as a method of reducing the ‘tax’ on our senses induced by an overload of information through the visual sense (Rosenberg, 1993). Milgram and Kishino, in outlining their taxonomy, recognise that it could serve to alleviate ‘analogous issues associated with other display modalities’, and cite work from Cohen, who at the time was developing realistic acoustic environment rendering in AR (1993). They also note the concurrent work in haptic (the various senses of touch) displays for augmented reality.



(a) Sutherland's 'Ultimate Display'

(b) Caudell and Mizell's 'HUDset'

(c) Feiner's 'KARMA' System

Figure 2.4: Historical AR HMDs

Mann terms 'Mediated Reality' (1994) as more of a method to computationally mediate our own perception, rather than just focusing on overlaying. His later work terms 'All Reality (*R)' which broadens the taxonomy of AR with concepts such as surveillance and privacy (2018). Azuma et al., acknowledge that AR is not just limited to visual perception or overlaying processes: 'AR can potentially apply to all senses, including hearing, touch, and smell. Certain AR applications also require removing real objects from the perceived environment, in addition to adding virtual objects'. With this in mind, what does the landscape of more contemporary AR design look and feel like today?

2.1.2 Forms of AR

Head-mounted

Successors of AR head-mounted displays (HMDs) devised by Sutherland (1968), Caudell and Mizell (1992), and Feiner (1993; 1997) exist today commercially within products such as the Microsoft Hololens 2¹, the Magic Leap ML-1², and the nReal Light³. These devices afford optical see-through capabilities through half-mirror visors and high-resolution displays, and offer stereo audio output. Input is managed mainly by hand-tracked interaction with overlaid visual menus and objects, but is also possible through eye tracking and voice control. The environment is sensed through a combination of 6DoF pose tracking, and spatial mesh mapping. For most HMDs today, this input and environment processing occurs off-device, on a tethered or wearable computer, or mobile device. These devices tend to be relatively expensive due to the extremely high R&D costs associated with the field (due to its novelty). Leap Motion's open-source Project North Star⁴ provides a much cheaper alternative at the trade-off of slightly larger form and lack of integrated computing and audio implementation, but has the highest field-of-view of all four and comparable tracking.

¹ <https://www.microsoft.com/en-us/hololens>

² <https://www.magicleap.com/magic-leap-1>

³ <https://nreal.ai/product/>

⁴ <https://docs.projectnorthstar.org/>



(a) Microsoft Hololens 2



(b) Magic Leap ML-1



(c) nReal Light



(d) Leap Motion Project North Star

Figure 2.5: Contemporary AR HMDs

A device that is head-mounted poses clear benefits to both input and output modalities. For input sensors, being mounted on the head provides them with more accurate tracking from the perspectives of our own senses (organic sensors). Consequently, and as Lindeman and Noma maintain in their classification scheme for multisensory AR, an important factor for the parity (if desired) of real and virtual output is a consideration of the location along the ‘stimulus pathway’ at which they mix: at the environment, sensory subsystem or computer level (2007). For a device to be head-mounted, allows ‘mixing’ to happen closer to the sensory subsystems, i.e. our eyes, ears, nose and mouth, thus resulting in an experience that is more closely intertwined with our real-time environment stimuli.

Handheld

Today, anyone who owns a smartphone is carrying an AR device in their pocket. This particular genre of AR development began in earnest in the mid 2000’s and today is the most widely used form of AR due to the pervasiveness of smartphones. It builds on the early research and practice of handheld AR applications built on cellphones and personal digital assistants (PDAs) in the late 1990’s, and smartphones in the late 2000’s, whose increasing variety and accuracy of onboard sensors: cameras, gyroscopes, accelerometers, and GPS, provided a convenient all-in-one solutions for developing handheld visual AR experiences.

The limitations of handheld displays are outlined by Bimber and Raskar (2005, pp. 79-83): reduced immersiveness due to the small screen size, compounded by the distance between it and the participant (normally arms length), impaired gestural input due to the need to hold the device; restrictions due to the cameras image quality and less computational power available to effectively register (align) real and virtual objects. As well as these issues, smartphones have developed to be primarily visual devices despite their name, and as such, their ability to deliver high fidelity audio, let alone audio AR mixing is limited due to their distance from the ear.

Recently, the lack of non-visual AR in handheld smartphones has been addressed through the increased adoption of ‘hearables’, a marketing term for small and unobtrusive, wireless earphones. More specifically, two features make these suitable audio solutions for AR use; ‘transparent’ hearing modes (see mic-through (Lindeman et al., 2008)), and integrated 6DoF⁵.

Projective

Projection of data into the environment in order to overlay or alter perception of real objects, i.e mixing virtual and real stimuli at the environment level (Lindeman and Noma, 2007), is perhaps the least adopted of Azuma’s categories of AR. The property that sets projective AR apart from head-mounted and handheld AR is perhaps its potential to deliver experience to multiple participants within a space. Although collaboration is possible through the previous two categories, it requires all participants to engage technology that is affixed to their body (either through wearing or through holding). Projective AR alleviates both the burden of physical attachment to a device, and also the expense of procuring a device for each participant. This category of display is typically realised by use of video projectors and speakers. Although the base requirement of registration in three-dimensions between virtual and real (Azuma’s third specification), is fulfilled relatively easily (add more projectors) compared to head-worn and handheld approaches (more sensor computation and accuracy),

Other forms

Of course, this categorisation is not exhaustive, and leaves out body worn AR systems such as the Audiomented Sound System (Chevalier and Kiefer, 2020), and tangible augmented

⁵ <https://www.apple.com/airpods-pro/specs/>

objects (Schraffenberger and van der Heide, 2015). This will be expanded on within my own framework towards designing multisensory AR instruments.

2.1.3 Sensory Display in AR

Within these common forms of AR technology, one could split the various technological processes that occur into input, process, output (the general IPO model found in software engineering), in AR this is often labelled: tracking, process, display. Tracking, as mentioned is often through hand gesture, body position and orientation and voice activation. Display is managed in the majority of cases through screens. The present section outlines various other sensory displays that have the potential of creating more immersive tools of expression.

Visual Sense

It is poignant to mention that until that last 20 years or so, sensory sciences, and as a result, technological displays of information, were focused on the visual, with the auditory following behind, and then the touch, smell and taste senses. In cross-modality psychology research, this ocularcentrism (the perceptual and epistemological bias ranking vision over other senses) is normally explained by a ‘textbook’ explanation: ‘the idea that vision is the most important modality is supported by numerous studies demonstrating visual dominance.’ Fabian Hutmacher argues that ocularcentrism can be critiqued through the lenses of (2019):

- A methodological-structural explanation: ‘Research on vision is often easier than research on other modalities and that this is the result of an initial bias toward vision that reinforces itself.’
- A cultural explanation: ‘The dominance of the visual is not a historical constant, but rather a result of the way (Western) societies are designed.’

However, AR is still typically realised in the form of graphical overlay, and as such, the forms which we find ourselves interacting with are predominantly based around screens and other forms of visual display (Dey et al., 2018). These are often split into optical and video see-through methods of display. The former employs semi-reflective half-mirrors to combine the reflection of close-by screen with the natural pass-through of the real world; the latter uses the feed of cameras to completely occlude the participant’s vision, and

performs AR processes on top of the camera feed. There are advantages and disadvantages to both methods, outlined in (Rolland and Fuchs, 2000). Head-mounted systems tend towards optical see-through, whereas most handheld AR (smartphones) are exhibit video see-through AR.

Auditory Sense

The second most developed-for sense in AR is the auditory system, borrowing from a rich history of spatial audio techniques such as exploiting HRTF for binaural audio playback and realistic audio localisation. Within the field of audio AR, as previously mentioned, ‘hearables’ with ‘transparency hearing’ modes offer the auditory equivalent of video see-through (coined mic-through (Lindeman et al., 2008)), where the occluded outside world is captured through a microphone, on top of which virtual sounds can be processed. Also outlined by Lindeman and Noma is the use of bone conduction headphones to offer a mediated perception of sound environments without the need to occlude the participants ears (coined hear-through), which could be considered equivalent to optical see-through. Moving into the environment, there are a myriad spatial audio techniques that could be used within AR, delivered by speaker arrays such as those found in wave-field synthesis and ambisonic beam-forming practices.

Haptic Sense

Due to the prevalence of hand-tracking technologies in the interaction with head-mounted AR applications, one might assume that being able to actually *feel* the virtual objects that are placed in AR would be one of the most important and developed concepts in AR. However, much of the early research in AR placed importance around overlaying instructions, rather than interactive objects. Hence, the haptic feedback of touching AR objects is relatively under-explored compared to the common issues which affected early and (funded) research - accurate registration of real and virtual processes; and higher visual fidelity. In most cases, it was deemed enough to *see* the effect that ones hands or body position had on the AR object or scene respectively. Today, there exist numerous technologies that could provide computationally mediated haptic perception in AR, for example through vibrotactile feedback or electrical muscle stimulation (Lopes et al., 2018).

The Chemical Senses

Whereas auditory and haptic sensations have been developed in AR, the chemical senses (smell and taste) have received far less exploration. Visual, auditory, and touch information can be believably recreated with technology through analog to digital conversion of electromagnetic radiation (light) and mechanical wave transmission (sound and touch). Conversely, smells and tastes contain organic information, the sensors (and thus A2D conversion) for which have not been invented yet. As such, the closest we can get to receiving chemical sensory data is through computationally activated rather than computationally mediated displays: scent emitters ([Maggioni et al., 2019](#)), and taste patches for example. Accordingly, in order to mediate the sense of taste, creative methods of sensory illusion have often been employed, i.e. Narumi et al. demonstrating that the sense of flavour can be mediated via visual overlay in AR ([Narumi et al., 2011](#)). The sense of temperature has also been mediated through cross-modal sensory illusion exploiting the trigeminal nerve in the nose, Brooks et al. demonstrates that users in VR can be made to feel warmer or colder based on the relative amounts of capsaicin or eucalyptus emitted near the nose ([2020](#)).

2.1.4 Computationally Mediated Realities: The AR Process

Additive layering of virtual content onto our real environment is by far the most typical of processes experienced in AR. However, it is only one of the potential methods of interaction between virtual and real components. In navigating this issue, the present thesis aligns its conception of AR closely with the definition of AR by Chevalier and Kiefer ([2020](#)): ‘real-time computationally mediated perception’, which fulfils Azuma’s three characteristics of AR systems, but is not as exclusionary as Caudell and Mizell’s original definition: a technology used to ‘augment the visual field of the user with information’. Investigating the existing relationships between real and virtual components of AR systems, Schrafenberger outlines at least five ‘relationships’ and five ‘subforms’ of AR (see [Table 2.1](#)). This taxonomy ([2018](#), pp. 80-130) provides an ideal starting point for those designing AR systems interested in creating embodied experiences that challenge the typicality of basic visual-overlay AR. Furthermore, Mann expands the ‘Reality - Virtuality Continuum’ ([Milgram and Kishino, 1994](#)) into the ‘All Reality Continuum’, integrating further dimensions such as ‘Metaveillance’, ‘Kineveillance’ and ‘Phenomenality’, such a standpoint not only ‘anticipates the need for an ethically aligned reality’ ([2018](#)), but demonstrates

Relationship	Description
Coexistence	Unrelated
Presence	Spatially Related
Information	Content-Based Relationship
Physical	Affect Each Other
Behavioural	Sense and React to Each Other
Subform	Description
Extended Reality	The Virtual Supplements the Real
Diminished Reality	The Virtual Removes the Real
Altered Reality	The Virtual Transforms the Real
Hybrid Reality	The Virtual Completes the Real
Extended Perception	Translating the Inperceptible

Table 2.1: Schraffenberger's taxonomy of relationships and emergent AR subforms

the multifaceted nature of augmented reality to afford experiences of ourselves, others, and environments computationally mediated by more than just additive informational overlays.

In particular, ‘diminished reality’, aims to *remove* real objects in our environment. When explored in the visual sense, this approach uses computer vision and content-aware substitution to remove a section or object of the real-world environment through camera sensors. Noise cancelling potentially offers an auditory diminished reality, but due to the fact that sound is propagated via mechanical waves, our sense of sound is closely intertwined with physically feelings of vibration, which are much harder to remove (Mori et al., 2017). Further examples of sound-driven but cross-modal techniques are outlined in (Walther-Hansen and Grimshaw-Aagaard, 2020).

‘Altered reality’ has the potential to provide new and otherwise inaccessible sensory and perceptual experience Schraffenberger describes that this is possible through virtual components altering the perception of real components in an AR system. Mann outlines a myriad examples of altered visual percepts (1994), from giants eyes, to ‘slow’ glasses. More recently, Nishida et al. have experimented with ‘egocentric smaller-person experience’ through use of a video see-through AR headset, in which unlike typical AR, the cameras are mounted at waist height, rather than on the HMD (2019). This afforded participants a difference in real-time visual perception, and resulted in them generally feeling smaller and behaving more like a child. In tandem with this altered visual perception, they propose a modified robotic glove ‘HandMorph’ (2020) that alters sensations of touch

through reduced finger and overall grip size.



Figure 2.6: Nishida's 'HandMorph'

In a similar vein, 'extended perception' revolves around the potential to provide *more* senses, or the sensing of stimuli outside of the range of our sensory systems. Bees for example have a visual system that extends far further into the ultraviolet range of EM radiation than our own eyes allow us; dogs hear much higher frequencies our ears allow us (thank goodness); and sharks can smell one part blood in a million parts water. These fun facts demonstrate vast differences in sensory acuity between species; but did you know that some shark have a constellation of pores⁶ along the lateral side of their bodies that allows them to perceive minute pressure changes in water? This enables them to create a 'pressure map' of their surroundings through active body movement. Modern technology could endow us with similar and myriad extended perceptions of our environment and its stimuli; Mann has experimented with making representations of radio waves translated down to the sensory range of our own eyes (2018)



Figure 2.7: Mann's 'Sequential Wave Imprinting Machine'

This section demonstrates the multitude of potential experiences that AR affords beyond mere visual information layering, next, I will examine the usage of AR in the arts, as a medium or tool for creating immersive aesthetic experiences, and demonstrate how this contributes towards a more holistic understanding of AR as a technology.

⁶ <https://www.sharktrust.org/shark-senses>

2.2 Augmented Reality in the Arts

Computational art, the use of computational programming or software as a tool, performance aid, medium or collaborator to create art or artistic works, is currently witnessing the adoption of new and exciting technologies such as trained machine learning algorithms, increasingly immersive displays and large-scale networked experiences. These implementations of technology within the creation artworks leads to a greater understanding of the materiality and affordance of these technologies. In this section, I draw attention to the modes of sensory engagement, aesthetic experience, collaborative expression, and activism that AR specifically affords artists willing to exploit its use as a medium.

2.2.1 Sensory Engagement

Computational art has seen the use of AR as a medium since the early 2000's: in 2008, Grasset et al. presented case studies of AR art exhibitions, with the conclusion that for effective design of artworks, specific importance should be placed on the relationship *between* real and virtual components (2008). These relationships have come to describe a matrix of possibilities between senses and mediating processes, i.e. altered hearing, extended smell, diminished sight. Papagiannis has drawn attention to this broader multisensory view of AR in attempting to understand an aesthetic for AR applications, 'AR is beginning to expand in new ways, beyond visual frames and into the full human sensorium' (2014). In 2015, the Tate Sensorium multisensory art exhibition used novel mid-air haptic devices to augment visual art (Vi et al., 2017). The results of the included study found that more congruent multisensory experiences lead to audiences finding artwork more emotionally engaging. If more congruent multisensory experiences of art lead to more emotional engagement, and AR has the potential to mediate multiple senses in rich and provocative ways, surely this makes it an ideal medium for such works.

2.2.2 Aesthetic Experience

Chevalier and Kiefer highlight the nascent use of newer AR technologies by artists (2020). They argue that AR has far more potential for creative exploration, and that it is a medium for creating 'new nuanced and fine-grained emergent aesthetic experiences' and as previously mentioned, necessarily define AR as 'real-time computationally mediated perception' in order to be inclusive of a multisensory approach to design.



Figure 2.8: Tate Sensorium Multisensory Art

The installation ‘Concrete Storm’⁷ by artists Gordijn and Nauta demonstrates the ability for AR to afford novel experiences of seemingly immutable and rigid physical objects (short concrete pillars), through the extension and animation of the concrete into larger, shifting and breaking structures. In describing the design of the work, they outline one of the main impetuses of creating the work: ‘Gordijn says the duo was interested in beginning to investigate the point at which viewers ‘might let go of trying to distinguish between what is real, what is not’ and accept this new mixed world as simply another version of reality’ (Gottschalk, 2017).



Figure 2.9: Concrete Storm

⁷ <https://www.studiodrift.com/concrete-storm-microsoft>

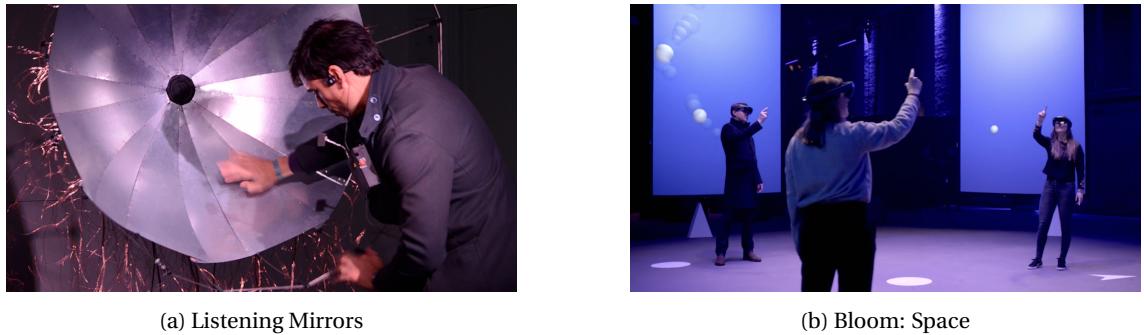


Figure 2.10: Collaborative Expression in AR

2.2.3 Collaborative Expression

Another key aspect of the use of AR as a medium in computational artworks is its ability to enable collaborative expression and construct shared perceptual spaces. For example, *Listening Mirrors*⁸, an artwork by Chevalier and Kiefer, (2018) is an AR installation that engenders a collaborative and performative AR sound environment. This environment is hybrid in nature; split between physical and virtual space. In the installation space, participants can interact with a audio augmented parabolic acoustic mirror. Their interactions are intertwined with a computationally mediated virtual sonic environment that they experience through bone conduction headphones. Additionally their vocalisations are mediated by a wearable microphone through the mirror. This invites collaborative expression through exploration of the shared sonic world.

Similarly, Eno and Chilvers explore collaborative ambient audiovisual composition in their AR adaptation of the popular app ‘Bloom’, ‘Bloom: Space’⁹. In this installation, participants wear a Microsoft Hololens and use hand gestures to activate and manipulate generative audiovisual elements. These elements can be experienced by other participants and thus create a shared perceptual and actionable space.

2.2.4 Activism and Agency

Due to its ability to be an effective tool for collaborative expression, or perhaps because it creates a liminal hybridity that ‘questions the possession and control of a physical space’ (Thiel, 2018), AR art has seen use in activism and collaborative political expression. Early pioneers such as the Manifest.AR group were able to create convincing and socially immersive AR art using visual layering¹⁰. They did this through aligning the collaborative

⁸ <http://listeningmirrors.net/>

⁹ <https://inculture.microsoft.com/musicxtech/bloom-open-space/>

¹⁰ <http://www.sndrv.nl/moma/>

potential of AR with political expression. By creating an app that overlaid visual art within the Museum of Modern Art (MoMA) through GPS, they allowed participants to visit an exhibition in the museum that had not been organised by MoMA themselves ([Figure 2.11a](#)). AR thereby afforded these artists the ability to embed virtual art within an area considered private property, and also promote collective AR agency, raising very early questions about ‘virtual trespass’ with AR. In a similar vein, ‘#arOCCUPYWALLSTREET’ (#arOWS) was an arm of the Occupy Wall Street protests of 2011. Some 25 artists succeeded in overlaying the Wall Street area with ‘over 400 protest related augments’, including audio overlays of protesters who were forbidden by police to enter the area. In this way, ‘AR was able to overcome their surveillance, barricades, horses, and excessive police numbers’ ([2018](#)) and more easily provide the means to action to people unable to physically mobilise.

As well as its liminality between real and virtual space promoting otherwise dangerous collaborative expression, AR has the power of rendering the invisible as seen, and due to this, it is a tool that has been used to shine light on social, economic, political and environmental injustices. Through its ability to change our perception of the real world through sensory augmentation, diminishment, hybridisation and extension, it has been used as a medium within the arts to ‘uncover’ underlying mechanisms in our society. For example, Thiel and PATTU’s (artists Cem Kozar and Işıl Ünal) 2011 work ‘Invisible Istanbul’ aims to render visible the unseen urban dynamics of Istanbul, through GPS positioning smartphone visual overlay. They write: ‘Viewers become as photographers: the act of viewing or making a screenshot of the objects at a specific site and time reifies the virtual objects into artworks, revealing hidden forces within the city not visible to the naked eye.’ ([2011](#)). In PATTU’s work, an augmented reality walking tour, different parts of the city are overlaid with symbolic information of past, present, and future uses of each area, highlighting tensions from military and commercial uses, and the effects this has on ‘contemporary urban space and the lives of its inhabitants’ ([2018](#)).

As summarised by Manifest.AR member Martin Skwarek puts it: ‘Alice has stepped through the looking glass [...] It is the job all future artists and activists to use this technology for the better, to bring people together, and uproot social injustice.’ ([2018](#))

2.2.5 An Arts Approach to Designing Multisensory AR

Agreeing that AR can be used as a medium for the composition of computational art through software tool or instrument design, surfaces the question of what the materiality, both for user and developer, of such a digital tool, or piece software would look, sound,



(a) Manifest.AR MoMA Invasion



(b) Invisible Istanbul

Figure 2.11: Activism through Interventionist ARt

feel, taste or smell like. As Papagiannis highlights, ‘Understanding the capacities of the technology and its constraints to exploit the technology to artistic use by envisioning novel applications and approaches, and developing new aesthetics and conventions beyond previous traditional forms’ (2017).

If visual overlay AR has been used as ‘a viewing instrument to bring into focus forces invisible to the naked or unknowing eye, and make them visible in the public sphere.’ (Thiel et al., 2011), what emergent properties might devices rising out from multisensory approaches to design afford? What are the futures of digital musical instruments, whose designers opt to use this technology?

The next chapter explores three theoretical approaches that will help us understand more about the hybrid real-virtual *space* it co-constructs with participants when in use, the *materiality* afforded by AR when used as a medium for creative expression, and the experience of *embodiment* it has the potential of imparting on these participants.

THEORETICAL FRAMEWORK

3

In actual experience, there is never any such isolated singular object or event; an object or event is always a special part, phase, or aspect, of an environing experienced world—a situation'

(Dewey, 1934, p. 67)

In the previous chapter, I outlined the increasing use of AR as a medium for the creation of meaningful experiences through interactive artwork. Examining the application of “real-time computationally mediated perception” (Kiefer and Chevalier, 2018) as a definition for AR, this chapter considers a trio of theoretical lenses through which to consider this process of mediation and the types of aesthetic experience this results in. Namely, I will consider the material, embodied, and spatial relations that AR fosters within this context of artistic production and experience.

3.1 Knowledge and Aesthetic Experience

What is experience when it comes to art; how does it relate to epistemological practice? To address this question, the present thesis draws from the work of the American pragmatist John Dewey (1859-1952). In his 1934 text, *Art as Experience*, Dewey argues that the production and consumption of art faces a crisis. After having been historically coupled in an constitutive relationship with the processes of everyday sociocultural life, it is being increasingly separated from these “conditions of origin and operation in experience” (Dewey, 1934), in a way that places art upon a remote pedestal. Dewey describes this as a product of the growth of capitalism. On proposing what they term Dewey Aesthetics, Leddy and Puolakka write:

“Nothing about machine production per se makes worker satisfaction impossible. It is private control of forces of production for private gain that impoverishes our lives. When art is merely the ‘beauty parlor of civilization,’ both art and civilization are insecure. We can only organize the proletariat into the social system via a revolution that affects the imagination and emotions of [hu]man[kind]. Art is not secure until the proletariat are free in their productive activity and until they can enjoy the fruits of their labor. To do this, the material of art should be drawn from all sources, and art should be accessible to all.”

(Leddy and Puolakka, 2021)

From this standpoint, art can serve as an emancipatory force for positive social change; but only on the condition that it is first brought back to the “origin and operation” of everyday experience — through the democratisation of a wider corpus of artistic media, tools, and social contexts in which these are deployed. In the 21st century, we could mistake this for already having happened. The increasing availability of ubiquitous technologies such as the internet, wearables, smartphones, powerful computers and software have shifted artistic production closer to the site of everyday sociocultural life — from the studio to the bedroom. Yet in doing so, has art really been knocked from its pedestal and been re-integrated into, and resituated to arise from everyday life?

Today, the fabric through which art tends to be disseminated and therefore consumed, social media, arguably operates within this same capitalist framework, the vocabulary having shifted from art in the museum, to content on our feeds. Despite making art more accessible to produce and consume, the capitalist logic of surplus extraction the centre of the algorithms that determine our interaction with social media vies to keeps art separate from everyday life. Central to this is that instead of being centralised in a museum or gallery, it is the concept of the pedestal itself that we should venerate. We must each curate our own museums and galleries, our own feeds.

The unavoidable truth of these platforms however, is that they are not solely focused on fostering individual or even collective curation of democratised art/content. They still employ the capitalist framework at their core. This extractivist profit motive defines the algorithmic fabric of online “content creation” (production), and resultant social media “engagement” (consumption) through mass data harvesting and advertisement selling. The longer users are engaged on a platform (Facebook, Instagram, Twitter, YouTube, TikTok), the more likely they are to generate profits for the platform via advert click/tap-throughs. Our feeds are interspersed with adverts and sponsored posts that have

been carefully curated to maximise the potential click/tap-through rate of their victims. Soshana Zuboff defines this as arising from a “market of human behavioural futures” [p?] in her book Surveillance Capitalism. Through this surveillance, large amounts of behavioural/affectual/personal data (termed behavioural surplus by Zuboff) are ‘skimmed’ off the top of our engagement with these platforms, and sold to agencies that match this personal data with products/content you are most likely to engage with. How could it be argued that the production and consumption of art / content on such platforms is not governed by “private gain” at our expense?

More recently, there are facets of the rise of NFT art projects that epitomise this continued fetishisation and veneration of digital fine art under the guise of “decentralisation”. Unfortunately enough, these projects often fall foul to the exact “centralisation” they market to oppose, creating small micro-communities of inter-centralised followers that collectively and actively, through social media, ensnare new and uninitiated retail investors via the urgency phenomenon of the “fear of missing out”. The motivation being to increase the value of their own NFTs. In most NFT projects, the only drive is profit motive, the digital art is merely the vehicle through which this investment is (im)materially realised. Despite the technological basis of some of these projects being quite radical in their proposition of democratic ownership and the privacy and security of digital information, the asset class most of them end up facilitating is tied to a pyramid scheme that necessitates the enrichment of the creators and early adopters of that specific NFT collection. How, in any way, could this method of production and consumption be said to be different from Dewey’s outlining of the rise of the “nouveaux riches”:

“The nouveaux riches, who are an important byproduct of the capitalist system, have felt especially bound to surround themselves with works of fine art which, being rare, are also costly. Generally speaking, the typical collector is the typical capitalist. For evidence of good standing in the realm of higher culture, he amasses paintings, statuary, and artistic bijoux, as his stocks and bonds certify to his standing in the economic world.” (1934, p. 7)

It could be argued therefore, that we have witnessed an increase in the amount of new media formats, but not one that resituates production outside of capitalist logic for the masses; even in so called “decentralised” projects. When Leddy and Puolakka stress that it is this “private control of [the] forces of production for private gain” that leads to this disconnection of art from experience, it follows that the creation of art outside of the confines of this control addresses the imperative for the drawing of art “from all sources”

[own emphasis]. From this view, the convergence of digital art to their “conditions of origin and operation in experience” could be seen as one of the primary foci of more recent movements such as Maker, Hacker, and DIY Labs, as well as the general ethos of participatory design and open-source hardware and software in art, design, and electronic music practice.

But what does it mean to say that art ought to originate and operate “in” experience, and how might these movements address this? Dewey considers the “live creature” in response to this question. He views the participant of aesthetic experience as meaningfully inseparable from the environment in which they are embedded:

“The senses are the organs through which the live creature participates directly in the on-goings of the world about [them]. In this participation the varied wonder and splendour of this world are made actual for [them] in the qualities [they experience]. This material cannot be opposed to action, for motor apparatus and “will” itself are the means by which this participation is carried on and directed. It cannot be opposed to “intellect”, for mind is the means by which participation is rendered fruitful through sense; by which meanings and values are extracted, retained, and put to further service in the intercourse of the live creature with [their] surroundings.” ([1934](#), p. 22)

From this basis, namely that perception, sensorimotor action, and intellect cannot be meaningfully separated when considering the intercourse of participants with their environment, Dewey highlights the weakness in the traditional dualist theory of mind - we are embodied and embedded beings. He posits that experience results from the “interaction of organism and environment”, and that in its fullest, experience can represent a “transformation of interaction into participation and communication”. It would follow, therefore, that in the production of artistic works, consideration of this dynamical relationship that constitutes our subjective experiences is not only valuable, but that artistic works should aim to arise (originate) from common and relatable states of this relationship. Returning art to this origination and operation “in” experience through interactive and participatory digital means can therefore be seen as a crucial mechanism for inducing positive societal change through fostering new channels of “participation and communication”. Shifting the production of artistic work away from the exploitative practices of mass manufacturing inherent in consumer technologies, as well as centring embodied participatory design practices, is a common theme among DIY and Makerspace communities.

This is not to in any way demean or reduce the social importance of art that does not take these approaches. It is also especially tricky to prove in any certainty the subjective measures of aesthetic experience that could result in positive societal change. However, I do view the imperative to shift production and consumption of art outside the capitalist architecture of current proprietary software tools and consumer technologies, and bringing the aesthetic closer to every day experience as a fruitful endeavour for the artist. I argue that doing so would emphasise the dynamical relationship participants find themselves in with their and, crucially, others' sociocultural, economic, and ecological contexts.

Where does knowledge lie for us as researchers in the production and consumption of such artistic works? On what basis can we test and evaluate these claims? One view that is popular in the field of music technology is that of considering the nature of the interactive and digital systems that we develop; a consideration of the epistemic tool.

“The digital instrument is an artefact primarily based on rational foundations, and, as a tool yielding hermeneutic relations, it is characterised by its origins in a specific culture. This portrayal highlights the strengthened responsibilities on the designers of digital tools, in terms of aesthetics and cultural influence, as they are more symbolic and of compositional pertinence than our physical tools.” ([Magnusson, 2009a](#), p. 335)

Developed by Magnusson, this line of thinking proposes that the digital instruments we design and employ as artists have inscribed in their physicality, affordance and sonic output, knowledge of cultural, historical, and designerly significance. In this way, they can be considered complex systems of research importance through the examination of their symbolic design, construction, performance, and appreciation by audiences. Magnusson embeds this proposition within Don Ihde's philosophy of 'instrumental realism' and Wittgensteinian philosophy, and draws from an understanding of the extended mind hypothesis. This theory, which will be expanded on in the next section, proposes that

“[Under certain conditions], the human organism is linked with an external entity in a two-way interaction, creating a coupled system that can be seen as a cognitive system in its own right.” ([Clark and Chalmers, 1998](#), p. 7)

Similar to Dewey's conception of how the live creature is inseparable from the environmental conditions that it is embedded in, the extended mind hypothesis and similar theories of embodied cognition prove to be invaluable in the field of tangible user interfaces, and digital music instrument design, due to the way they can explicate as well as

draw out the dynamic relationships between artist, collaborator, material, interface, and audience in the deployment of expressive artistic media. In the following sections, I shall dive deeper into these theories of mind (ToM), and what they have to offer a practice that makes use of augmented reality technologies within the field of computational art and musical performance.

3.2 The 4E Approach to Experience

One way of expanding on this conception of aesthetic experience is to adopt the 4E approach to cognition and experience. Key to my following theoretical underpinning of materiality, embodiment, and space will involve how this approaches the themes of perception, agency and cognition arising through the complex interthreaded coupling between a participant and their environment. Often referred to as 4E cognition (4EC), it states that cognition is an embodied, embedded, enactive, and extended process (Gallagher, 2017). It's worth mentioning that 4EC is an amalgam of related theories of mind (ToM) across the disciplines of philosophy and the social and cognitive sciences, and as such there are different flavours of it. Many of these have said to have roots in the Dewey's pragmatism (among others') outlined in the previous section, as well as Merleau-Ponty's embodiment theories (Zavota, 2016). What these constituent theories all bear in common is the rejection of the standard cognitivist model of experience that states that cognitive processes happen solely 'in the mind' and 'in the brain'. Another principle is that they generally draw from complex systems theory: "understanding the complex interplay of brain, body and world requires the tools and methods of nonlinear dynamical systems theory" (Clark, 1999). Within this theory there are a multitude of phenomena that can explicate the interactions between components of such a continually unfolding system, including (De Domenico and Sayama, 2019):

- Interactions - a complex system is formed of many components, many of which will be connected in a network of mutual and on-going interaction
- Emergence - from this network of interactions, there exists the possibility for complex and novel structures to form that are inexplicable from the features of their simpler components
- Dynamics - the state of a complex system can change over time, often in an unpredictable non-linear fashion

- Self-organisation - the components of a complex system may exhibit collective behaviours, and perceived large scale structure
- Adaptation - complex systems don't necessarily shift between steady states, they actively and reactively respond to environmental stimuli
- Feedback - structures of a complex system exhibit a phenomena where the outputs of a system are routed back in, via closed chains of cause and effect

4EC normally begins with an acknowledgement of the complex embodied nature of human cognition. Varela, Thompson, and Rosch propose that cognition is embodied in a way not too dissimilar to the assertion Dewey holds in regards to the “senses” of the “live creature”.

“By using the term embodied we mean to highlight two points: first, that cognition depends upon the kinds of experience that come from having a body with various sensorimotor capacities, and second, that these individual sensorimotor capacities are themselves embedded in a more encompassing biological, psychological, and cultural context.” (1993, pp. 172-173)

Alva Nöe's concept of sensorimotor contingencies has since built on the first point of this approach to embodiment, in a way that proposes an “enactive approach to perception” accepts that perception is something we do, not something that happens to us (Noë, 2004). This has since been termed the ‘sensorimotor approach’ to enactivism, and Gallagher notes that it falls short of an “enactivist approach” due to its omission of affective aspects of embodiment such as “mood-related and emotional factors, [...] bodily states such as hunger, fatigue, and pain, as well as a complex motivational dimension that animates body-world interaction” (Gallagher, 2017, p. 150) The proposal is that that an ‘enactivist’ approach, i.e. one that aligns with Varela, Thompson, and Rosch's claim that sensorimotor capacity is “embedded in biological, psychological, and cultural contexts” more holistic perspective, the proponents of embodiment theories of 4EC propose a radical shift from the standard cognitivist ToM; sensory perception and bodily existence is not only causally related to the emergence of cognitive processes, it literally constitutes them.

A natural continuation extends this line of reasoning to the material environment in which the participant is a part of— accepting that it is in turn shaped by the web of sociocultural norms and values that Dewey alludes to, and is the site of sensorimotor action. Mark Rowlands proposes this underlying thesis as “the embedded mind”:

“In accomplishing cognitive tasks, an organism can utilize structures in its environment in such a way that the amount of internal processing it must

perform is reduced. Some of the complexity of the task is, thereby, off-loaded onto the environment, given that the organism has the ability to appropriately exploit that environment.” (2010, p. 68)

It necessarily follows that if a cognitive system that has the potential for environmentally embedded sensorimotor action, i.e. it is embodied and embedded, it also enacts its cognition. The first formal account of what could be called an enactive approach to cognition begins also in “The Embodied Mind”, by Varela, Thompson and Rosch. From a foundation of accepting that sensory and motor processes, namely perception and action, are “inseparable in lived cognition”, enaction is contingent on two assumptions: perception consists in perceptually guided action, and cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided (1993, p. 173). The authors argue that an enactive approach to cognition therefore proposes that perception must be studied from the point of view of the participant’s sensorimotor structure, rather than any “pregiven, perceiver-dependent world”, i.e. the world is not perceived through an internal representational model of what is “outside”, rather, the world emerges from a participants ability to “guide [their] actions in [their] local situation”, to enact.

The last assertion of a 4EC account claims that cognition has the potential to be extended into objects in a participants environment. The extended mind hypothesis, as developed by Andy Clark and David Chalmers and described in the previous section, is the core thesis of this proposal. The authors propose that, cognitive processes, beliefs for example, can be “constituted partly by features in the environment, when those features play the right sort of role in driving cognitive processes” (1998, p. 12). They compare the experience of cognition in two different hypothetical people, Inga and Otto. Both are tasked with recalling the address of a museum and travelling to it. Inga has a neurotypically functioning memory, and so accesses the memory of the address (which assumedly lay dormant previous to this action) recalls it, and travels to the address. Otto has Alzheimers, and like many others with Alzheimers relies on the use of a notebook or memory aid to recall experiences and other important information, he finds the address in the notebook, and travels to the museum. In both cases, Clark and Chalmers propose that the cognitive functionality of what constitutes ‘belief’ are analogous:

“To say that the beliefs disappear when the notebook is filed away seems to miss the big picture in just the same way as saying that Inga’s beliefs disappear as soon as she is [no] longer conscious of them. In both cases the information

is reliably there when needed, available to consciousness and available to guide action, just the way that we expect a belief to be" (1998, 13).

It therefore follows, that cognition is extended, beyond the boundary of the mind-body and into the environment, and even into specific artefacts - a key theory that enables the proposition of Magnusson's epistemic tool within the field of music technology, as mentioned previously. Taking these four points of departure, namely that cognition is embodied, embedded, enacted, and extended, Gallagher presents the following assumptions as the core conditions of an enactivist approach to cognition (Gallagher, 2017, p. 6):

- a) Cognition is not simply a brain event. It emerges from processes distributed across brain – body – environment. The mind is embodied; from a first-person perspective embodiment is equivalent to the phenomenological concept of the lived body. From a third-person perspective the organism–environment is taken as the explanatory unit.
- b) The world (meaning, intentionality) is not pre-given or predefined, but is structured by cognition and action
- c) Cognitive processes acquire meaning in part by their role in the context of action, rather than through a representational mapping or replicated internal model of the world
- d) Enactivist approaches have strong links to dynamical systems theory, emphasising the relevance of dynamical coupling and coordination across brain – body – environment
- e) In contrast to classic cognitive science, which is often characterised by methodological individualism with a focus on internal mechanisms, enactivist approaches emphasize the extended, intersubjective, and socially situated nature of cognitive systems. Enactivism aims to ground higher and more complex cognitive functions not only in sensorimotor coordination, but also in affective and autonomic aspects of the full body.
- f) Higher-order cognitive functions, such as reflective thinking or deliberation, are exercises of skilful know-how and are usually coupled with situated and embodied actions

Poignantly, this way of framing our human experience is salient for the digital humanities at this point in time. More of our everyday social, artistic, political, cultural, environmental, and economic interactions are mediated by technologies and platforms. These

are privately owned by mega-corporations, that as previously mentioned, have shown dubious a relationship with ethics concerning misinformation and user privacy, as well as extractivist methods of guaranteeing shareholder profit. Does Facebook's 'metaverse' sound like the kind of AR environment that sounds like a safe and democratic ground for art that seeks the "transformation of interaction into participation and communication" (Dewey, 1934)? This decision is up to the reader, but I would argue not as it will become clear in the following sections.

Out of this model of explicating experience, I propose three lenses through which to examine AR applications in the sonic arts - materiality, embodiment, and spatiality. In some ways, a linear, written account of these processes, separated neatly by subheadings is somewhat incompatible with the logic of 4EC and its process-inseparability, nevertheless, in the following sections of this chapter I will attempt to go into detail how we can use this model of experience, to consider the themes of materiality, embodiment, and space within AR musical instrument research practice, as well as the broader study of digital humanities. Loosely, these three sections will cover the artistic, audience, and societal considerations and potentialities respectively, when engaging with AR systems.

"The image with which the artist works to realise his or her idea is no longer a phantom, it can be touched, navigated and negotiated with."

(Ryan, 1991, p.5)

3.3 Augmented Materiality: The Relational Fabric of the AR(tistic) Medium

Thus far, we have outlined the contextual origins of AR technology, as well as their contemporary physical forms, modes of sensory display, and methods of real and virtual processing. Additionally, I have outlined the theories of aesthetic and perceptual experience through which these technologies could be interrogated to further understand AR's role as a medium for artistic expression. In this section I will propose a lens through which to consider the artist's composition of 'material' in AR experiences as a method of fostering a modulated dialogue between elements of a participants sensorimotor engagement with their perceived real and virtual environment.

3.3.1 Complex Material in Musical Interfaces

Existing frameworks for considering the design of relationships between real and virtual processes do exist within in the field of AR; in the early 90's, the theoretical work of Milgram and Kishino, their reality-virtuality continuum (1994, p. 10), proved to be a fruitful way of taxonomising the types of AR and VR technologies that were arising from different fields at the time. However, for my purposes of detailing the material nature of an artists' composition in AR, a more nuanced and detailed framework is needed, one that goes beyond specifying the what (a spectrum of reality to virtuality); towards investigating of the how, (a processual space of augmentation, diminishment, alteration, hybridisation, and extension through which real and virtual environments are enmeshed), and the why (the conveyance of meaning and intentionality of the artistic work).

Hanna Schraffenberger, as highlighted in [Section 2.1.4](#), offers a set of relations that describe the dynamics between "the virtual and the real" (2018). She argues that from the fundamental relationships of presence, information, physicality, and behaviour, there emerge new "subforms" of AR, namely extended, diminished, altered and hybrid reality, as well as extended perception. Yet, even with this level of description, the processes themselves are still abstract - these are descriptors for the types of processes, but what are the processes themselves? I believe that the answer lies again, in an approach aligned with complex systems theory; through a consideration of phenomena such as interaction, emergence, dynamics, self-organisation, adaptation, and feedback.

Within the field of music interfaces, considerable research and practice has been conducted in the area of complex systems. Surmised neatly by musical interface researcher Joel Ryan, hidden, or black-boxed complex processes within digital interfaces can be developed and explored through ascribing "physical handles" to these "phantom models", e.g. providing embodied interaction to algorithmic musical processes to better understand, design, and evaluate them.

"The physicality of the performance interface helps give definition to the modelling process itself. The physical relation to a model stimulates the imagination and enables the elaboration of the model using spatial and physical metaphors. The image with which the artist works to realise his or her idea is no longer a phantom, it can be touched, navigated and negotiated with." (1991, p.5)

Perhaps the design of AR musical instruments is concerned with not only providing physical handles to phantom models, but phantom handles to physical models too.

In the same vein, Agostino Di Scipio outlined his perspective on the paradigm of interaction in the context of audio signal processing and musical interface use.

"Interactive music systems are dedicated computational tools capable of reacting in real time, upon changes in their 'external conditions', [typically including] initial input data and run-time control data [...] set, changed and adjusted by some agent – a performer, or group of performers [...] Control devices, with their mechanical and/or visual interfaces, are operated to determine these data.

The main purpose of control data is to determine a system's changes of internal state. This is done indirectly, by updating the parameters in either digital signal processing techniques or program routines operating at a more abstract, symbolic level. Changes of internal state are heard as changes in the musical output. By operating the available control devices, the agent in effect 'plays' the system as if it were a new kind of music instrument" (2003, p. 1)

For Di Scipio, the act of integrating the agent / performer into the definition of the music system implies that 'interaction' exists in the "underlying ontology" or "interdependent coupling" of participant and instrument; at the site of their "interface". This coupling creates linear feedback loops in performance, between the resultant sound, and decisions made by the performer. Indeed, this lower-level dependency (human-machine) itself exists within the wider "meta-system" including "[human], machine and environment". Within this meta-system, later referred to as an "ecosystem", Di Scipio outlines that:

"A principal aim would be to create a dynamical system exhibiting an adaptive behaviour to the surrounding external conditions, and capable to interfere with the external conditions themselves. Not only would it be able to detect changes in the external world and 'hear' what happens out there, [...] it would also be able to become a self-observing system, that is, to determine its own internal states based on the available information on the external conditions – including the traces of its own existence left in the surroundings. [...] This [move towards composing musical interactions] should be described as a shift from creating wanted sounds via interactive means, towards creating wanted interactions having audible traces. In the latter case, one designs, implements and maintains a network of connected components whose emergent behaviour in sound one calls music." (2003, p. 6)

The depth of research based on this type of complex interactive music system (IMS) has deepened considerably in the last 20 years, as artificial intelligence and machine learning algorithms, low-latency microcontrollers, and real-time compositional software have been implemented into what Magnusson terms “intelligent instruments” (2009b, p. 8). Thus, “performative ecosystems” (Waters, 2007) such as these exhibit behaviours of complex systems as mentioned above, including interaction, emergence, dynamics, self-organisation, adaptation, and feedback. I propose therefore that an enactivist approach to experience (4EC), one that posits the emergence of cognitive processes across the distributed system of mind-body-environment as being wholly appropriate in the design, creation, performance, an evaluation of interactive music ecosystems.

Some of the first proponents of this approach include Essl and O’Modhrain, who propose an enactive approach they term “weak sensorimotor integration” to the design philosophy of tangible interfaces for musical expression. “What concerns us here[...], is the consideration of enactive knowledge in the context of musical instrument design and how perceptually guided action defines the ‘feel’ and playability of a musical instrument.” (2006, p. 3). Core to their research is the question: “In what way does a specific integration of sensory perception and motor action correspond to a controllable, even enjoyable musical performance experience?” Building on the class of expected sounds that correlate to our action with physical objects in the world, Essl and O’Modhrain evaluate the design of a set of prototype tangible interfaces that introduce a flexibility in this expectation-realisation coupling between action and sound by altering the digital sonic output of the system based on measured user interaction. They propose that the resultant interaction with these instruments provided grounds to include a loose sensorimotor integration into the design of musical interfaces. The present thesis is interested in the implication of this hypothesis in the creation of not only physical interactions with IMS, but also virtual, and ultimately, hybrid interactions. My own 4EC approach to the compositional material of AR musical experiences draws from Essl and O’Modhrain’s own motivation for their proposal of an enactive approach. How are dynamics of AR musical systems learnt over time? How is the ability to predict musical outcomes affected by the integration real-time motion sensing and parameter mapping? How does the configuration of hybrid sensory content define the way action becomes perceptually guided?

Newton Armstrong also provides considerable insight on the enactive design of DMIs. He proposes an “enactive model of interaction” that is concerned with “circular chains of embodied interdependency between performer and instrument, instrumental “resistance”

to human action and intentionality, and an integrative approach to the roles of sensing, acting and cognitive process in the incremental acquisition of performative skill." For Armstrong, this approach draws from a principled distinction between 'functional' and 'realizational' interfaces. Whereas a functional interface "serves a predetermined function; it is structured around a finite set of interactions which are known in advance of the task's execution", a realizational interface "brings with it the possibility of continuously realizing new encounters and uses, and, in the process, of re-determining the relationship between technical objects and their human subjects" (Armstrong, 2006, p. v). This distinction, and a pursuit of the latter over the former, means to prioritise the contexts of "meaning and signification in which human and medium are embedded, and is conducive to dynamic and indeterminate forms of interaction" (p.42). These dynamic forms of interaction are realised through performer intention, and instrumental "resistance" in response the performer's engagement with it. Over time, this resistance establishes adaptation in participant interaction, constituting what Armstrong terms a "dynamical responsiveness" between the performer and their instrument (pp.46-47).

Lauren Sarah Hayes' work approaches live electronic music practice from an enactive and embodied perspective. This work often uses tangible and haptic controllers that provide non-linear controller input to various elements of the "ecological network of sound analysis and digital signal processing" that constitutes her performance ecosystems. For Hayes, designing complex and adaptive musical systems allows "elements of instability, vulnerability, unpredictability [to] become fodder for the improviser" (Hayes, 2018, p. 2). Hayes' creative and research practice has enabled a pedagogical methodology, in which "music and movement-based improvisational practices" are explored in classes of undergraduate students. Through a combination of deep listening methods, movement exercises, and free improvisational exercises, Hayes argues that there is a potential for co-creation of compelling improvisational performances in classes of up to thirty untrained improvisers. These techniques result not only in the active exploration of musical sensorimotor contingencies, but also in the creation of musical meaning that is "brought forth through an emergent extended cognitive network that involves complex relationships between the improvisers involved, technology, and the space in which the performer-instrument pairings are distributed" (Hayes, 2019, p. 8). These ensemble improvisations draw from a wide variety of media formats including AR. Pointing towards the following assertion by Schiavio and van der Schyff (2018), Hayes notes that digital musical instrument practice ought to develop and augment improvisational practices "that build

collectively on individual musical and creative sociocultural histories that individuals bring to the table”:

“If the body plays a key role in determining musical learning, so does the socio-material and cultural environment in which it is embedded. However, these dimensions are not separated from the body. Rather, they become manifest through the body and are co-determined by actions and interactions with other bodies and things in socio-culturally meaningful contexts.” (Hayes, 2018, p. 9)

Chevalier and Kiefer propose a similar method of promoting musical expression in their audio AR (AAR) work “Listening Mirrors” (2018). Their system is “audience participation dependent”, and could constitute a realisational DMI. Through real-time analysis and re-projection of participants’ breath and vocal sounds via bone-conduction headphones and a transducer-augmented parabolic mirror, audience members, environment, and interfaces “feed off each other”, co-creating physical and virtual sound environments that overlap and co-construct a shared hybrid listening space. Their creative and research practice is situated in the later work of Merleau-Ponty and related theories of enactivism. In particular, the authors draw from Merleau-Ponty’s notion of the “flesh”:

“In defining what is meant by ‘flesh’, Merleau-Ponty states, ‘[w]e must seek space and its content together’, that we are ‘interwoven into a single fabric’, a ‘universal flesh’, and ‘he who sees cannot possess the visible unless he is possessed by it, unless he is of it’. The notion of ‘flesh’, therefore, is both the ‘flesh of the world’ and the ‘flesh of the body’, the relation of the corps sauvage and cultural world and its representations.” (Chevalier and Kiefer, 2018 citing Merleau-Ponty, 1945; 1968)

Augmented reality, specifically due to its ability to mediate perception in real-time and across a multitude of sensory modalities, presents the opportunity to “experiment with, modulate and disrupt [conditions of situated-ness, timeliness, emergence, multimodality and engagement inherent to our embodied coupling] to create new audience collective experiences” (Chevalier and Kiefer, 2018)

Thus, depending on the specific artwork, from any or each of the DMI designer’s (Essl and O’Modhrain, 2006; Armstrong, 2006), improvisational performer’s (Hayes, 2019), or installation artist’s (Chevalier and Kiefer, 2018) perspective, if the material composition of the system with which they are engaging is designed to provide a realisational, dynamic interaction then it “assumes open-ended, fluid, and at least partly indeterminate processes

of signification, and as such requires the ongoing cognitive involvement of the musician" ([Armstrong, 2006](#), p. 48). This necessity of engagement aesthetically situates such artworks and instruments closer to the site of everyday embodied life, what Dewey terms the return of art to its "origin and operation" in experience. These examples also align themselves with the 4EC approach to experience, which argues that cognitive process are embodied, embedded, enacted, and extended, across a distribution of mind - body - environment.

Core to my own approach is, as I'm sure is evident by now, is the centring of the idea of the processual nature of AR systems. What are the processes; when I say real / virtual processes what is it that I am trying to convey? Ocularcentrism and the additive layering paradigm have shifted discussion of AR systems towards an object-centred view of the "reality" that they present, this effect is also present in VR ([Hovhannisyan et al., 2019](#)). Instead, I argue that an AR object is not a static visual thing with marked boundaries, not in any meaningful way to the artist/designer at least. It must be thought of as an on-going and distributed component existing inside a dynamic web of experience and meaning; as such, any discussion of the contents or material of an AR composition ought to be considered through this lens. Even a single apparent 'virtual object' that is in front of a participant is not *just* that; it is an invitation, a handle, a real-time process by which a participant can be perceptually guided through sensorimotor action towards a specific aesthetic experience. This so called 'object' might be part of a larger organised whole, a component of a room-scale musical experience for example, or an emergent property of a hidden set of complex conditions that have only just been met through the specific movements of a participants body. This, more holistic, view of AR processes allows for a more fruitful discussion of the types of experiences we can expect to craft as AR practitioners; and engages more critically with Schraffenberger's taxonomy of AR relationships.

3.3.2 The Medium Specificity of AR

Embedding Schraffenberger's fundamental relationships ([Table 3.1](#)) within Di Scipio's notion of an "ecosystemic" musical interface, or Water's "performance ecosystem", i.e. viewing the AR process as a real-time network of interactions distributed across the 'hybrid' brain - body - environment, helps us in identifying the medium specificity of AR, and how it may uniquely be exploited to create meaningful AR artwork. If the motive for the composition of an artwork is to intentionally impart knowledge, meaning, or truth through aesthetic experience these must exist within and across the above network of interactions

Relationship	Description
Coexistence	Unrelated
Presence	Spatially Related
Information	Content-Based Relationship
Physical	Affect Each Other
Behavioural	Sense and React to Each Other

Table 3.1: Schraffenberger's Fundamental Relationships

and relationships in our brain - body - environment distribution. I suggest we view this latent aesthetic experience as an real-time interaction between the components of:

- Intention - the intended meaning behind the composition of the piece
- Medium - the specificity of the apparatus by which this meaning is imparted
- Experience - the on-going process by which the medium is engaged with by the participant
- Realisation - the resultant knowledge, truth, or subjective experience from of the above

In the first section of this chapter, we proposed that from Dewey's perspective, an artwork ought also to originate and operate within everyday sociocultural life in order to bring about positive social change through aesthetic experience. These concerns could be conceived as mainly arising from the component of intention, and its dynamic interaction through the medium into experience. Concerning the physical, or indeed virtual, manifestation of these intentions is that of the material composition - which is facilitated by the medium and its materiality. Of course, a consideration of the composition of a work cannot be divorced from the real-time experience of a participant, but from the perspective of the artist wanting to engage in these tools, an understanding of the inherent nature of the medium of AR is necessary. Realisation of specific messages in the experience of artistic works could manifest in various ways, but one could argue that "the aesthetic experience" of the artwork has the potential to incur a feedback loop in which the participant is in a constant state of realisation, due to the nested, non-linear, and explorative aspects of the instrument or experience - drawing from Armstrong's realisational versus functional interface. Perhaps they even come away from the artwork, imparted with changed beliefs or outlooks on the content of the experience; and then goes on to act on these beliefs within their sociocultural life, thus leading to others potentially changing their own beliefs too.

In relation to the above constituent parts, in the context of Schraffenberger's relationships, what defines AR's medium specificity when chosen for the creation of expressive works of art, and how can these relations provide fertile ground for a new aesthetic of composition? The key element of these relations I argue, is the underlying assertion that what makes AR unique is its ability to modulate the perceived conceptual 'distance' between real and virtual elements in three-dimensions and in real time. In Azuma's ([Azuma, 1997](#)) original definition of AR, this is referred to as the 'registration' or 'alignment' of virtual content to real world content. I argue that Schraffenberger's fundamental relationships expand this concept beyond just a spatial alignment of elements. Presence-based relationships modulate this spatial distance between elements, while Information-based relationships modulate the thematic distance, Physical relationships modulate the material distance, and Behaviour-based relationships modulate the ecological distance.

The narrative around the modulation of these various conceptual distances between real and virtual processes inevitably tends towards the closing of the gap between them. After all, the industries developing AR view this as the 'issue' of registration - the virtual and the real must be brought closer together. For example, the proximity of a headset to a participant's eyes, as well as the resolution and acuity of motion and tracking sensors, brings virtual objects closer or more aligned to the perceived reality of the physical environment of the participant - thus closing the spatial distance through a Presence-based relationship between that 'virtual object' and the physical environment. Another example might be an audio-tour. In this example, an Information-based relationship is invoked by closing the 'thematic distance' between virtual and real components, here, an abstract informational audio script becomes embedded in the actual environmental content of its real world setting. Schraffenberger argues that real and virtual objects casting shadows on each other constitutes a physical relationship - I'd argue that it is key to also view it as a reduction in the perceived material distance between the two objects (the way in which their physical matter behaves in respect to each other converges on the expected outcome of if both objects were physically real). Also proposed is that a virtual animal reacting to real world sounds would constitute a behavioural relationship - as you might expect, I'd argue that it must also be viewed as constituting a reduction in the perceived ecological distance between the animal and the sound (the set of behaviours that constitutes the interrelation of both environments converge on the outcome that would be expected if both animal and sound were members of our physical reality). These discussions of what I term 'closing the gap' usually fall under the banner of the drive for "increased immersion"

or “believability”.

However, just as AR is demonstrated to be able to close the spatial, thematic, material, and ecological distance between the virtual and the real, so too can it further the gap between them. This is what Schraffenberger encapsulates in her argument for the proposal of the ‘relationship between the real and the virtual’ to replace the proposal of the ‘registration of the virtual to the real’ by Azuma among others. What if the thematic distance between the real and virtual is radically increased, but all others are kept the same? Could this be exploited for aesthetic effect? For artists using AR, this is an important consideration — along with their definition of “real”: is it synonymous with “truthful”, “physical”, “tangible”? What is the resultant aesthetic experience of participants if the virtual content of the artistic or musical AR scene is divorced from the expectation of how a physical counterpart engages with space, theme, matter, and ecology? What about when going beyond representations or remediations of existing physical objects and processes, and presenting participants with radically novel virtual processes that still seem to be “embedded” in our physical reality via these technologies?

From this line of reasoning, the medium specificity of AR could be said to be its “invocation of a performance ecosystem constituted of relationships between real and virtual processes in the axes of spatial, thematic, material and ecological distance”. If these relations are in turn experienced by a participant whose cognitive processes are embodied, embedded, enacted, and extended, it may stand to reason that the closer the distance between physical and virtual elements on these axes, the less discernible “physicality” and “virtuality” may become, and the cloudier the boundary between them. Despite being slightly alarming, this isn’t to make the argument that a “virtual piano” might ever be mistaken for a “physical piano”, but I believe that the claim could be made that given enough time, a participants notion of what is “real about a piano” has the potential to be modulated, given that the virtual piano is at some level altered to be incongruous with its physical counterpart, but on most axes indiscernible from it. Perhaps it may look the same, but behaves differently, i.e. the keys play from high to low when pressed right to left. In this way, for a participant, the meaning and concept of a piano could change through this broader process of sensory or perceptual illusion, and then have real consequences in the physical world. This is a fairly benign example compared to what may be possible with future AR technologies, and I view this as an important consideration for artists to hold: what are the ethical considerations I need to make as an artist who is using technologies that enable experiences such as this? What platforms am I using, and is

telemetry gathered for a corporation of the technology that I am using? If AR experiences are asking participants to suspend their disbelief in exchange for new realities and beliefs, as artists we must be clear on what these beliefs are, and how they might be realised after the experience, in the actions of participants. This is explored further in the section of the chapter on space.

As artists and musicians, I would argue that a focus on providing multisensory engagement in an AR work be of paramount importance in most of the above considerations. This not only provides more channels through which to modulate the distances described found in Schraffenbergers relationships, but also widens the broader ecosystem of interactions possible for an enactive participant of their hybrid environment. This is nature of human experience proposed by 4EC - that the participants of an AR experience are embodied beings that cognise through perceptually guided action that is in turn afforded by their sensory paraphernalia. Schraffenberger and van der Heide provide useful insight here. They propose that multisensory AR is the "norm rather than the exception", due to the fact that physical reality is already multisensory, despite the ocularcentrism found in AR devices. They propose three routes through which to pursue this kind of design. Firstly, integrating non-visual sensory displays into AR systems. Secondly, taking seriously the assertion that "the real world plays a crucial role in the resulting experience" (Schraffenberger and van der Heide, 2016, p. 5). This is to say that physicality and virtuality become entwined in a set of dynamical relations that must be considered as "more" than their constituent elements. Chevalier and Kiefer argue that this highlights that AR is "inseparable from a multisensory ecosystem, inhabited by modes of sensing, modes of perceptual mediation, computational relationships between sensing and mediation, human participants and their environment" (2020, p. 4). This gestalt, you could refer to it as, when viewed through my own ecosystem approach to AR constitutes the relational distances of spatial, thematic content in AR. Thirdly, they claim that the nature of a visual 'virtual object' can become more believable in our perception of it if it behaves and is affected by non-visual real-world environmental stimuli - such as the virtual creature visually responding to external sounds - this has the effect of closing the material and ecological distance despite not providing virtual non-visual stimuli itself.

In her taxonomy, Schraffenberger describes these as Presence-based, Information-based, Physical, and Behavioural relationships. These fundamental relationships underlying the experience of an AR participant could be said to have been founded in the artist's intention to construct a performance ecosystem that contains varying conceptual distances between

real and virtual processes in the axes of space, theme, material, and ecology. However, these relationships exist outside of the realm of in-the-moment agency for a participant, in some kind of representational or intentional belief-system of the artist. They are not the actual handles by which a participant perceptually guides their actions. Instead, those affordances are based in the hybrid real/virtual reality that emerges from said relationships or performance ecosystem. Schraffenberger describes these as “AR subforms”. These processes of reality modulation are what could be seen as the ideal ground through which to bring art back to the origin and operation of everyday life. But where does this leave artists interested in taking advantage of this medium specificity?

If we take seriously this proposed model of AR — that the resultant experience of participants is one that constructs a complex system of simultaneously real yet virtually modulated, subverted, augmented, or diminished hybrid environments — it follows that these environments provide a hybridity of options for new modes of perceptually guided action. Moreover, taking 4EC as a basis for understanding an audiences experience of such dynamic relations, and defining augmented reality as “real-time computationally mediated perception” ([Chevalier and Kiefer, 2020](#)), it follows that AR has the ability to afford novel modes of aesthetic experience that affect our cognition. Through this intertwining of real and virtual processes in the enactive space of participants, AR presents an opportunity to uniquely render the typically invisible, unheard, and intangible tensions and injustices in our everyday cultural, socio-economic, and environmental realities. For the artist, AR offers itself as a novel medium for such creative work - these tensions and injustices having long been one of the central narratives of artistic production.

*If Martians land in Washington
Square Park, they won't see the
virtual piano.*

(Chalmers, 2022, p. 229)

3.4 Augmented Embodiment - Aesthetic Experiences of Embodied AR Systems

We started this chapter with an examination of John Dewey's account of the separation of art from everyday life, and the effect on it thereof. To account for the notion of an aesthetic experience in the appreciation of artistic works, Dewey points towards the 'live creature'. This concept, as discussed, contributed to the modern study of 4EC: the notion that cognitive processes ought to be thought of as embodied, embedded, enactive, and extended across the distribution of brain-body-environment. The previous section outlined my approach to material composition from an artists perspective. The medium of AR may be thought of as an "invocation of a performance ecosystem constituted of relationships between real and virtual processes in the axes of spatial, thematic, material and ecological distance". The present section moves forward to consider the experience of participants in such artworks.

3.4.1 Embodiment and Enactivism in VR

Considerable amounts of research related to embodiment in the context of virtual environments (VEs) generally falls under specific but necessary discussions and measurements of particular cognitive and affective processes such as presence, emotion, awareness, and consciousness (Slater et al., 1994; Seth et al., 2012), for example, in body swap experiments (Slater et al., 2010), the rubber-hand illusion (Suzuki et al., 2013), and stimulating altered visual phenomenology (Suzuki et al., 2017). Integration of explicitly 4EC approaches to VEs/VR have only started to appear in research in the last 10 years, and is generally lacking from the discourse around AR. For example, through an two and a half year auto-ethnographic exploration of her own experience in the popular VE Second Life, Maeva Veerapen integrated the phenomenological theories of Merleau-Ponty and Husserl to analyse and explicate the specific location of 'embodiment' inherent in the connection to her virtual avatar. She surmised that "such an experience is still anchored within the

user's physical body and that its relation with the avatar's body creates the link between the two spaces, physical and virtual" and termed this state of being symbembodiment - symbiotic embodiment of a combined real and virtual self (Veerapen, 2011). Drawing on Veerapen's notion of symbembodiment and linking it to an enactivist approach (indeed, as Varela, Thompson, and Rosch state in their "The Embodied Mind", Merleau-Ponty and Husserl's theories form part of the Western phenomenological roots of their approach (1993, pp. 173, 18), Willans et al. argue that the "episodes of emotion" in the experience of VEs are linked with feelings of 'presence', which in turn emerges from the embedded nature of a participants sensorimotor structure, or, we might say, enactment (2016, p. 23). Hovhannisyan et al. argue for an enactivist approach to understanding participant flow, which is motivated by the authors' resistance to the physicalist approach to designing "object-centred" VR experiences that is typically employed (2019, p. 1). In particular, they turn attention towards what they call an action-predicated perspective (one that focuses on the enactivist theory of perceptually guided action) that they state is "non-reductive with respect to subjective experience", and honours "the pragmatic dimension of perceptual reality" (2019, p. 18).

3.4.2 Embodiment and Enactivism in AR

Within AR, Riva et al. propose that AR and VR can transform our "external experience" through the increased levels of presence and engagement they afford, that in turn generates "personal efficacy and self-reflectiveness" (2016, p. 10), they further propose that it has the potential to do the same for our "internal experience". For the authors, "bodily self-consciousness" can be enhanced and extended by altering/extending its boundaries through a process called 'augmented embodiment', which builds on the hypothesis that embodiment could be altered, expanded, and distributed through AR technology:

"Whenever computer-based information is blended with the perception of the surrounding physical world, as in augmented reality this may become integrated into a new form of altered embodiment. But that requires that the augmentation of the physical with the virtual be carried out in such a way that the user has the potential to feel present. Given the clear popularity of mobility and social connectivity, it seems that presence will increasingly be experienced through attention to an external world in which the physical and the virtual are somehow blended" (Waterworth and Waterworth, 2014)

From an 4EC approach, “feeling present” in the context of bodily ”mobility and social connectivity”, emerges, and could be seen to be somewhat analogous with, the notion of action (and other cognitive processes) being embedded in the socio-cultural and -material norms and values of a participant’s environment, their inherent sensorimotor embodiment, and as a result, the propensity for their actions to be perceptually guided (enaction). Suzuki et al. operationalise Nöe’s (2004) concept of sensorimotor contingencies, as further developed by Seth in the field of predictive processing (2014), by using AR/VR technology as a testbed for exploring the effect of modulated sensorimotor contingency and congruency in participant’s visual experience of familiar and unfamiliar 3D objects (Suzuki et al., 2019). Their results indicate that “that the contingency, but not the congruency, of a person’s actions and their visual consequences influences access to visual awareness”, providing support for a 4E-like perspective on perception (and also hinting at the ability for AR/VR technology to be used as a method for purposeful sensory illusion through sensorimotor incongruence). Most recently, Chalmers has argued that AR has the potential to extend our mind in a more seamless fashion, in a way that he likens to Heidegger’s proposal of the “ready-to-hand” nature of certain tools; in use they become an extension of our body without our having to ‘think’ about them and the actions they afford.

“What sort of mental processes do augmented reality devices extend? For a start, they can more seamlessly extend all the processes that smartphones extend: memory (remembering someone’s birthday), navigation (getting to the museum), decision-making (deciding where to eat), communication (talking to friends), language processing (translating from another language), and more. But their immersive connection to our perceptual system provides new avenues for extension” (Chalmers, 2022, p. 299)

Chalmers further proposes that AR provides the grounds for further cognitive extension: our abilities of perception (e.g. through infra-red sensing), recognition (through algorithmic processing and machine learning of real-time environmental data), and imagination (through real-time replacement or transformation of physical objects with virtual ones).

3.4.3 From Relations to Realities

In searching for the medium specificity of AR through a close examination of Schraffenberger’s fundamental relationships, I argued in the previous section that core to the specificity

Subform	Description
Extended Reality	The Virtual Supplements the Real
Diminished Reality	The Virtual Removes the Real
Altered Reality	The Virtual Transforms the Real
Hybrid Reality	The Virtual Completes the Real
Extended Perception	Translating the Inperceptible

Table 3.2: Schraffenberger's AR subforms

of AR as a material for composition lies an artist's un/intentional weaving of conceptual distances between real and virtual processes in the axes of space, theme, material, and ecology. These construct realities, proposed by Schraffenberger as "AR subforms", and extend the definition of AR - towards a system that encompasses a variety of hybrid (real/virtual) processes that occur in experience. Most of these subforms, all of which were outlined in [Section 2.1.4](#), have seen nascent use in AR applications, not least in the arts. Due to consumer AR devices mostly taking the form of a headset or screen with either a camera feedthrough or optical reflection techniques (visual see-through or optical see-through), most applications fall into the category of the literal "augmenting" of reality, that is, to add to reality. However, as I have outlined and exampled, there are other modes of perceptual modulation that necessarily fall under a broader and more holistic AR definition such as Chevalier and Kiefer's "real-time computationally mediated perception" ([2020](#)).

The AR subforms in question ([Table 3.2](#), deal in the addition, removal, transformation, completion, and translation of environmental aspects through which a performer or participant's actions could be perceptually guided — an ecosystem of hybrid processes that are spatially (localised), thematically (contextually relevant), materially (are in/congruently tangible), or ecologically (reactive to environmental cues) embedded within their physical environment.

[Author Note: 6 concept-illustrations will be here of multisensory AR experiences, one for each subform and with captions]

What happens in these processes? What do they invite or necessitate a participant to think, do, and believe? Drawing more analytical depth from the six main assumptions of a 4EC approach to experience, we can ask the following questions in order to develop direction for further understanding of the nature of the network of a participant or performer, an instrument or experience, and their environment:

- In what way do the artistic choices that produce different AR subforms, e.g. hy-

brid reality, diminished reality, result in different modulations, i.e. varied initial conditions for the emergence, of cognitive processes, e.g. learning, beliefs, affect, expectation, memory?

- If the world is structured by cognition and action, which are in turn perceptually guided, in what ways do different AR subforms promote actions that disrupt, draw attention to, or re-structure this environment?
- Can specific AR subforms provide aesthetic experiences that disrupt the notion of representation mapping, internal models, and the standard cognitivist model?
- How can disruption of brain-body-environment coupling be operationalised to promote the emergence of new (inter)subjective and culturally situated musical meaning?
- How do the different subforms specifically facilitate embodied knowledge, as it pertains to the performance or experience of artistic works?
- Do the specificities of particular subforms of AR provide varied propensities for higher-order cognitive functions, e.g. does diminished reality lead to specific instrumental know-how?

From a 4EC approach, these above situations provide us with an interesting question: if the means by which aspects of the physical world is sensed can be obfuscated to an extent that removes said aspect, in what meaningful way can the participant's action in relation to it be perceptually guided anymore? What about in cases wherein sensory mediation results in the transformation or 'completion' of an aspect; in what ways does a participant's action potential change; how does this affect higher-order cognitive functions such as sense-making in relation to it? Does this lend credence to the notion that specific AR processes have the ability to construct new realities, rather than just provide illusions over the top of existing ones (Chalmers, 2022, p. 230)? If a participant of an AR musical experience, or a performer of an AR musical instrument engages with this plurality of perceptual mediations, might the addition, removal, transformation, and completion of aspects of the environment have the potential to alter their own embodiment, relations to their environment, and their enactive potential in specific areas of a space? If so, this particular method demonstrates a radical method of disrupting, or provoking new self-organised states between the real and virtual nodes of such an embedded performance ecosystem.

"Our environment is not a play, a performance in which only minor modifications are possible. It's an improvisation,[...] the nature of the interaction is radically open."

(Vermeulen, 2015)

3.5 Augmented Space - The Construction of AR Hybrid Spaces, and Avoiding The Snow Crash

As we have established so far, augmented reality, when applied as a medium for the creation, performance, and expression of multisensory art and music, has the potential to:

- a) Provide new modes of instrumental performance and expression to performers and participants
- b) Radically modulate the embodied experience of performers and participants
- c) Scaffold new musical and artistic meaning through these augmented aesthetic experiences

It follows that such experiences appropriate existing physically real spaces to do so, whilst either (a) simultaneously creating a relational virtual space dependent on the content of the experience, or (b) appropriating physically real spaces to the ends of constructing an entirely new hybrid AR space that is greater than the sum of its physical and virtual counterparts. Lev Manovich, terms this “augmented space”, and defines it as “the physical space overlaid with dynamically changing information [...] likely to be in multimedia form and often localised for each user” (Manovich, 2006, p. 2). The material and embodied nature of nascent XR technologies provide a new urgency to the consideration of what these spaces will be like to inhabit; who will construct them; to what extent they rely on surveillance technologies; who holds the keys with regards to software and hardware production. In the following section, this space will be referred to as either Manovich’s augmented space or, synonymously, AR hybrid space; both should be read as giving emphasis to the ecosystemic nature of AR’s real and virtual processes and subforms thus far highlighted in the chapter. In current popular culture, this notion of a type of physical and virtual hybrid space has been given the term (the) “Metaverse”.

3.5.1 The Metaverse

“The Metaverse” gained popularity as a term partway through this writing thesis, partly due to Facebook’s calculated re-brand to Meta, partly due to the resulting ‘bull’ run in “metaverse” cryptocurrencies, and partly due to the large subsequent increase of venture capital funding in XR start-ups. The reality is however that it is not a new term, having originated from Neal Stephenson’s *Snow Crash* (Stephenson, 1992), a sci-fi novel detailing an anarcho-capitalist future in which a VR-based internet (The Metaverse) has become the gamified site of socio-cultural and economic norms, values, and interactions. In *Snow Crash*, Metaverse users immerse themselves via personal or publicly available terminals (see VR), although using public terminals has a negative affect on your avatar appearance and thus your status in The Metaverse.

Though perhaps tangential to, rather than aligned with, the motivations and political thought contained in Stephenson’s work, *Snow Crash*, and other novels by him, have influenced Silicon Valley tech oligarchs such as Bill Gates (Microsoft), Jeff Bezos (Amazon), Sergey Brin (Google), John Carmack (Oculus, now owned by Facebook), and Peter Thiel (PayPal, and first external funder of Facebook) (Rogers, 2021). One of the auxiliary aims of this chapter is to ask, “with regards to the unfolding corporate metaverse, is theirs a morally just pursuit?” mainly through the question “should we be willing immersants of their techno(dys/u)topian dream?”. Broadly speaking, the definition of “the” Metaverse has since expanded to contain any sufficiently “immersive web3.0” interaction or XR experience. Most of these amalgamated definitions are still based in techno(dys/u)utopian ideals of the cybernetic man. This isn’t surprising; the technologies the Metaverse makes use of (AR, VR, GPS) lay “deep within the military and Western - scientific - industrial - patriarchal complex” Davies (2004). A result of this is a paradigm that is also found in *Snow Crash*: the unregulated (see neoliberalism) pursuit of acquiring or constructing virtual “land” or commodities only then to rent/sell it to either smaller corporations, or directly to the individual’s drive for commodity accumulation through property and material ownership. This is a needless relationship and is already enacted in the present-day “metaverse”, and it is helped by the appropriation of XR development and experiences as a “use case” or “value proposition” for Blockchain and cryptocurrency:

“[Living in the metaverse means] having a place to call home, where you can show off your possessions and maybe even have friends over” (own emphasis)
(Marr, 2022)

As mentioned in Knowledge and Aesthetic Experience, cryptocurrency, blockchain technology, and NFTs are the current landscape on which more broadly digital art is being appropriated to increase the price of cryptocurrencies. Artworks, in-game item collectibles, and limited-edition music albums; it is this specific method of “decentralised” property ownership and wealth procurement is the current platform on which a war is being waged on aesthetic experience for the betterment of the bottom line of tech oligarchs, and early cryptocurrency adopters. Stephenson’s “Metaverse” is already here, vis-à-vis virtual land ownership; many argue it started with Second Life in 2003, but today, cryptocurrency projects like Decentraland, Sandbox, and Somnium Space, as well as corporate efforts like Meta Horizon Worlds, Workrooms, and Home (by Facebook) are the logical conclusions of those early progenitor virtual environments, when it is taken into consideration the lack of regulation around "Big-tech" and cryptocurrency platforms in general:

“Republic Realm paid a record \$4.3 million for land in the largest metaverse real estate platform, Sandbox. The company is developing 100 islands, called Fantasy Islands, with their own villas and a related market of boats and jet skis. Ninety of the islands sold on the first day for \$15,000 each and some are now listed for resale for more than \$100,000.” ([Frank, 2022](#))

This article also mentions that despite \$500,000,000 having been spent in 2021 on virtual land on platforms like The Sandbox, average land prices have decreased 91% over the last 9 months since their peak in early December 2021, which occurred 1 month after Republic Realm’s purchase of \$4.3 million ([Kane, 2022](#)). How much of that “property” was bought by or let to retail investors hoping to become rich off of fiat currency savings, only to see 91% depreciation in their virtual asset?

All of these platforms essentially offer the same thing: a virtual “space” that is supposedly decentralised and ungoverned, where you can purchase virtual “land” that you can make your own, and that has a function either now or in the future. If that is the case, why are there so many different metaverse platforms? Is there a difference in the functionality of the land they provide? Is some more arable than others; more predisposed to host enhanced social interactions? To some extent, yes, certain platforms provide gaming experiences, others art curation and others still music experiences. Unfortunately, they are all ultimately built atop of the same type of scheme that is used by NFT projects described earlier on in this chapter. Developers of such systems employ artificial scarcity (core to most cryptocurrency-based environments) and thus the phenomenon of the “fear

of missing out", in order to 'encourage' market interaction and increase the monetary value tied to their own assets.

Taking a look at one of the most popular interactive "Metaverse" platforms, Axie Infinity (AI), which brands itself as a "play-to-earn" (PTE) gaming metaverse, we see exactly the type of emergent behaviours that come from its initial set of conditions, i.e. intentional artificial scarcity, low levels of government regulation on cryptocurrencies. In AI, players "breed, battle, collect, raise, and build kingdoms for their Axies" - think Axolotl-like Pokémon that are themselves NFTs ([Sky Mavis, 2022](#)). Until April 2022, players had to own 3 Axies before being able to take part in gameplay, Axies only being purchasable from breeders through cryptocurrency. The cost for a starter team peaked at a fiat conversion of around \$600-800 just to start playing the game. Upon winning player-versus-player battles with your Axie team, you earn in-game currency (at a fluctuating market price, dependent on a healthy and constant influx of new players), and thus fiat currency through conversion.

The gameplay of AI has been criticised as basic, repetitive, and boring by its own players. Indeed, qualitative analysis of player sentiment found that 57% (n=1713) of players polled found AI enjoyable, and only 44% (n=1313) reported making any financial return ([Delic and Delfabbro, 2022](#)). One emergent characteristic of the AI community (and PTE games in general) is that of 'Scholarship' programmes. In these, wealthier players dubbed 'managers' (or even just cryptocurrency investors and NFT holders — "payers" is increasingly becoming synonymous with "players"), rent out their Axies to prospective players who cannot afford the entry fee for a team of Axies ('scholars'). In return, scholars play the game daily for the manager, until a specific in-game currency quota is reached, then provide managers with a percentage (often between 50% - 80%) of that daily income. This results in a type of gamified labour exploitation, wherein the time of players from economically poorer countries is valued at less than, and exploited for, the profit of those in wealthier countries. This all whilst income (for the scholar) and profit (for the manager) is predicated on the necessity for increasing influxes of new players. This mechanism was reported on at the time, mostly in a positive light for the "economic freedom" it was offering younger players, especially in the Global South, in countries such as the Philippines ([Puyat, 2021](#)), and Venezuela ([Ledesma, 2021](#)). Research from several Filipino academic institutions (40% of the player-base resides in the Philippines) have provided more well-rounded critiques. One study on the lived experiences of 20 Filipino AI players indicated that whilst playing the game has lead to positive effects such as increased income, and the

creation of new communities during the COVID-19 pandemic, it has also brought about stress and lack of sleep due to the long hours of play needed to meet quotas (De Jesus et al., 2022).

But promises of economic freedom, artificial scarcity of goods in order to inflate demand and prices, dependence of growth through the investments of ever increasing numbers of players, and an unofficial recruitment scheme that allows for the exploitation of labour for profit of those higher up in the game? That does sound awfully similar to an unregulated corporate pyramid scheme. However, it wasn't until the cryptocurrency markets took a downhill turn (December 2021 - September 2022) just after their peak, where the effect of this was really felt for players of Axie Infinity. This was compounded by a high-profile hack that drained \$650 million from the platform. Following the resultant 91% depreciation in market prices, the time investment for scholars started looking even more abysmal in comparison to their income (which was tied to the in-game currency value). This resulted in a lot of scholars leaving the platform, much to the anger and entitled cries of the managers and guilds that thought they were under their employ, (Ongweso Jr., 2022; Gach, 2022)

What I hope to highlight in this section is that the “metaverse” is not what it is portrayed as. Its current co-option by (or perhaps origination in) the profit motive inherent to cryptocurrency-based metaverse platforms does not “liberate art”, rather it leads to exploitative, and hyper-commercialised practices. My criticism of these platforms is not to say that a diversity of media experiences, hosting a multiplicity of sociocultural narratives, from a variety of voices isn't welcome; rather the opposite. XR technologies, AR specifically, due to its material and embodied relations, propose a real potential to provide a real type of embodied artistic liberation. But as Leddy and Puolakka note on Dewey Aesthetics [Section 3.1](#), this liberation will never happen under the exploitative constraints of the capitalist modus operandi. Thus, “space” in "the Metaverse" will always mirror these conditions and be inherently flawed, yet if we are to believe the marketing speak, it is in ‘spaces’ like these that technology companies foresee cultural production shifting to in the next 10 years (Fatemi, 2022). What affect will this have on the production, consumption, and dissemination of art of all types?

3.5.2 Avoiding the Snow Crash

From a digital humanities perspective, the promised definition of “The Metaverse” (see blending of physical and virtual human existence) is really only a rehash of already estab-

lished philosophies on technological embodiment and theories of space, but plated up with generous lashings of neoliberalism, and served fresh to attract new venture capital. As such it can be readily (perhaps cynically) dismissed as a co-opted marketing buzzword that has little meaning and relation to the concepts, technologies, and labour that are ushered under its umbrella for the sake of technological progress, here synonymous with capitalist growth (see also Internet of Things, Artificial Intelligence, Machine Learning, 5G Networking, Blockchain). These philosophical grounds will be explored in the following section, where I set out a perspective on how AR's use as a medium for computational art and music could lead to the co-construction of new, hybrid spaces wherein novel, existing, hidden and suppressed realities can be acted out anew.

This section is not concerned with the creation of, or contribution to "the Metaverse", rather, it views novel art and music composition with AR a fertile ground for the staging of tactics, using Michel de Certeau's phrase, to "unsettle and diverge from the conventions" (1984, p. 36) of such heavily consumerist spaces. By placing importance on the embodied experience, and through creating novel and interactive aesthetic realities, the artist, musician, designer fosters strong and dynamic links between participants or performers and their environment - co-constructing socially and / or culturally significant 'augmented space'.

De Certeau makes a clear distinction between place (*lieu*) and space (*espace*): "space is a practiced place" (1984, p. 117), and embeds this within a linguistic context, "in relation to place, space is like the word when it's spoken, that is, when it is caught in the ambiguity of an actualization", "stories thus carry out a labor that instantly transforms places into spaces or spaces into places" (p.118). For de Certeau, place is constituted in the calculations of those with "will and power" (often institutions) to exert and isolate it (what he calls strategy), and as such, place "implies an indication of stability". In contradistinction, space exists within the "polyvalent unity of conflictual programs or contractual proximities", and is "actuated by the ensemble of movements deployed within it" and is thus subversive and operational; the result of "calculated action determined by the absence of a proper locus" - or "tactics".

Taking a comparatively more phenomenological perspective, and therefore perhaps more aligned with a 4EC approach to the spatially embedded nature of lived experience, Henri Lefebvre argues that "space is socially constructed", and that it is an "emergence" of the "social and mental" that is "produced" (1991, p. 260). Lefebvre argues that space "embraces a multitude of intersections, each with its assigned location" (p. 33) these are named in

his conceptual triad:

Spaces	Forms	Modalities	Equivalents
1 st	Spatial practice [<i>pratique spatiale</i>]	perceived [<i>espace perçu</i>]	subjective real everyday live/nature
2 nd	Representations of space [<i>représentations de l'espace</i>]	conceived [<i>espace conçu</i>]	objective imaginary urbanism/cartography
3 rd	Representational spaces [<i>espaces de représentation</i>]	lived [<i>espace vécu</i>]	collective symbolic lifeworld/culture

Figure 3.1: 'Lefebvre's Triad of Spaces in (Günzel, 2019, p. 14)'

Thus, de Certeau's distinction of space from place, whilst valuable insofar as it indicates the imbalance between human actors and institutions within the socio-cultural context - "the street, defined by urban planning is transformed into a space by walkers", in itself is hardly a novel realisation. Furthermore, its fecundity diminishes with the assertion that this relation should be understood solely through a linguistic lens, i.e. the rules of "place" are written by those with power, and "space" is the result of actions by participants reading and interpreting those rules. To view the relation as such is to say that the actions of participants must within the diktat of place, and thus asserts that space originates in place and is therefore inseparable - "the only freedom you have is to formulate alternative sentences" (Vermeulen, 2015).

As Vermeulen further notes, for de Certeau, "space is an inter-subjective activation of a static site, a place", whilst for Lefebvre, "place is the momentary suspension of a social flow, [a] space". This leads to a contradiction in the formulation of "agency" as a concept: "de Certeau understands agency as the enactment of a script not our own, whereas Lefebvre sees it not as a container for action but as the construction of action itself."

It's clear that with respect to 4EC approach to *live* aesthetic experience it may be difficult to reconcile the 2nd and 3rd spaces of Lefebvre's triad (conceived and lived), due to its own anti-representationalist standpoint. However oppositional on that front however, Lefebvre's argument, through the prominence of social and class struggle under capitalism underpinning its formation, does indeed align itself with certain components (embodied, embedded) of enactivism: an approach that emphasises "the extended, intersubjective, and socially situated nature of cognitive systems" (Gallagher, 2017, p. 6):

"The relationship to space of a 'subject' who is a member of a group or society implies a certain relationship to [their] body and vice versa." (Lefebvre, 1991, p.

Additionally, Lefebvre's "own particular brand of Marxism which stressed the importance of everyday life" (Merrifield, 1993, p. 8) could be seen as aligned with Dewey's own assertion of the importance of artwork to return "origin and operation" in everyday experience. In this way, the construction of 2nd and 3rd space doesn't necessarily have to fall fully under the remit of 4EC to explicate in real-time or live experience, rather, they may constitute the socio-cultural, and environmental aspects (norms, values, laws, traditions, and conditions), as well as higher-order cognitive functions (meaning, and interpretation) that form the 'environment', as it is then to be experienced *or* later considered by the enactive cogniser.

Thus, whereas "the Metaverse" is somewhere to "show off your possessions", and "maybe even" stage social interaction (Marr, 2022), augmented space (or AR hybrid space) in contrast, prioritises the social, cultural, and aesthetic experience of the everyday. Architecture in "the Metaverse", due to its origin in a capitalistic conception of society and technology, alienates its own inhabitants, since "under capitalism, it is only only through that [market] mediation that humans interact with buildings at all" (Miéville, 1998, p. 18). China Miéville's critique here of physical architecture under capitalism could also be applied to the profit motivation for the commodity fetishisation of art in NFTs and the Metaverse. Not just 'virtual land', but digital art, music, video games, and social connection itself. More broadly, his "Marxist phenomenology" argues that to focus entirely on the physicality (of architecture in his text) is to "ignore the profound experiential ramifications" of living in a social system where such commodities are exchanged for profit.

In stark contrast, architecture in augmented space consists in a constantly unfolding dialogue between humans and their environment, as perceptually guided by an ecosystem of hybrid processes; rather than consisting in alienation, and the "anxiety" induced by the commodity fetishism underlying capitalist profit motives. Application of Lefebvrian spatial theory indicates that AR can be a real force for social change, due to the way it can intervene spatially in the perception of participants, leading to new spatial characteristics and possibilites. When this is taken into action by people, either audiences, or performers, it puts the decision of when and where to "momentarily suspend" the social flow of space – and thus construct "place" – in their hands and in their bodies.

When Manifest.AR virtually trespassed the Museum of Modern Art (MoMA) with invisible and intangible AR art (Veenhof and Skwarek, 2010); when #OccupyAR broadcast the disembodied voices of geographically separated activists into Wall Street (Skwarek, 2018); when Cem Kozar and Işıl Ünal (2011; 2018) revealed the unseen urban dynamics of the city

of Istanbul, it was through the live co-construction of new hybrid spaces — augmented spaces — that once hidden realities could be enacted anew. Whilst one might see parallels between these actions and what de Certeau terms “tactics”, the former, however, do more than just subverting place ‘proper’ through “alternative sentences” to borrow Vermeulen’s phrase. In actuality it is far more radical. Through the dynamic relations present in its hybridity — its incessant movement — emergent and novel states of self-organisation can occur within its constituents part(icipant)s. New socio-cultural meaning through sensory, perceptual and environmental modulation have the potential to synthesise from this ongoing process. In that spatiotemporal ‘place’ then, neither the MoMA nor Wall Street were themselves.

3.6 Conclusion

This chapter started with a grounding in the concept of ‘art as experience’, drawing from the work of (Dewey, 1934). Dewey states that the capitalistic reverence of fine art, notably by the "nouveau riches", has led to art being put on a pedestal - disconnecting it from its origin and operation in the everyday experiences of people. Leddy and Puolakka have since proposed that "material of art should be from all sources and art should be accessible to all" (2021). Dewey points towards the aesthetic experience that art entails, and notes that the "live creature" is inseparable from the environment in which they are embedded. Today, despite social media as a platform for cultural production of artistic works has some merits, the fabric on which it is built is no different from the capitalist mechanisms that separated art from experience in the first place. Thus, open-source, hacker, DIY, and maker approaches were proposed as a method of democratising "all sources" of art, and attempting to increase accessibility. The focus on the body by Dewey also points towards the move towards tangible interface development, participatory design, and human-centered design.

To delve deeper into this experience of art, I propose an enactivist approach to considering cognition in experience. Often called 4E cognition (4EC), it states that cognitive processes are embodied, embedded, enactive, and extended (Gallagher, 2017). In considering the design, performance, and experience of AR, 4EC offers a fresh perspective on concepts such as agency, action, perception, immersion, and space.

The first theoretical lens through which to examine AR’s design in computational art and music practice involved outlined the importance of a complex systems framing

of interface use and performance. I drew on existing research from the field of digital musical instrument design (Magnusson, 2009a; Di Scipio, 2003; Essl and O'Modhrain, 2006; Armstrong, 2006; Hayes, 2019; Chevalier and Kiefer, 2018) in order to demonstrate the usefulness of considering performance systems as complex, and ecosystemic. Combining this theory with Schraffenberger's taxonomy of relationships between the virtual and real in AR, (2018), I developed the notion of "axes of conceptual distance", a loose framework by which to think about the (un)intentional design space that AR affords. The design of AR experiences could therefore be thought of as "invocation of a performance ecosystem constituted of relationships between real and virtual processes in the axes of spatial, thematic, material and ecological distance". The section is ended on the assertion that despite these concepts, and Schraffenberger's relationship types being useful for the artist, they don't describe what happens in experience.

To examine what happens in the experience of participants, the second lens was one that took seriously the assertions of 4EC, namely that cognitive processes in experience are embodied, embedded, enactive and extended. I draw from multiple disciplines of similar theoretical and practical work in VR and AR to show how 4EC offers a novel framing by which to consider the experience of AR by participants, audiences, and performers. I refer back to Schraffenberger (2018), and outline multisensory examples of her "AR Subforms", and propose a set of key questions that may bring about new understandings of aesthetic experience in AR; e.g. *"If a participant of an AR musical experience, or a performer of an AR musical instrument engages with this plurality of perceptual mediations, might the addition, removal, transformation, and completion of aspects of the environment have the potential to alter their own embodiment, relations to their environment, and enactive potential in specific areas of a space?"*

Closing out the trio of theoretical lenses, is a grounding of the aforementioned consequence of AR for the experience and construction of "space". I first critically examine the term "Metaverse" (Stephenson, 1992). Its modern-day origin is in technologies like AR and VR, which in turn sprung from "deep within the military and Western - scientific - industrial - patriarchal complex" Davies (2004), and its current operation has been heavily co-opted by cryptocurrency projects. Because of this, in its current state, the "Metaverse" is hardly fertile ground for the "live creature" to experience art "accessibly", and from "all sources" (Dewey, 1934; Leddy and Puolakka, 2021), indeed, it mirrors the exact profit and exploitation motive of the historical context it originated in. I use the term "augmented space" (Manovich, 2006) to develop the proposition that R's use as a medium for compu-

tational art and music could lead to the co-construction of new, hybrid spaces wherein novel, existing, hidden and suppressed realities can be acted out anew, and ground this in the social space theory of Henri Lefebvre (1991), and of contemporary examples of AR activism (Veenhof and Skwarek, 2010; Thiel et al., 2011; Thiel, 2018; Skwarek, 2018).

In the next chapter I will outline the methodologies taken within my research, and draw links between historical and contemporary uses of AR, and the theoretical lenses I have proposed here.

A DIY APPROACH TO AR IN THE ARTS

4

This chapter draws some of its content from a published short-paper in the Doctoral Consortium of the Tangible, Embedded, and Embodied Interaction Conference 2021

(Bilbow, 2021b)

4.1 Research Methods

The methods present in this thesis have stayed consistent with their original proposal, with the exception of specific methods of data collection and analysis being disrupted by the COVID-19 pandemic, which commenced three months after the start of research. In Chapter 2, the historical origins, contextual trends, exceptions, and contemporary forms of AR technologies were outlined. Additionally, examples of AR's use in the arts were provided, along with specific rationales for their use, e.g. sensory engagement, collaborative expression, and activism and agency. In the previous chapter, I outlined a trio of lenses through which to consider the material, embodied, and spatial aspects of AR. These drew from a 4EC approach to experience, which posits that cognitive processes are embodied, embedded, enacted and extended. In the current chapter I will be outlining my methodological approach to the practice-based side of the research: designing, performing, and evaluating musical AR experiences, and AR DMIs. A review of the outstanding research questions will be outlined, to provide the chapter a clear direction into the ensuing three study chapters.

-
3. What are AR's boundaries and affordances as a medium, are there problems it poses artists and musicians?
 4. What is the resultant experience for participants and audiences (or 'immersants') of an artistic AR experience?
 5. What might a future corpus of artistic AR experiences and digital music instruments look, sound, and feel like?

The present thesis approaches the above questions by making use of several established practice research methods. Broadly, these methods have lead to practical contributions that will be explored in this chapter and the following three study chapters: namely the creation and iterative design of multisensory AR experiences, and a set of open design patterns for doing so. These practical contributions function both as ongoing research probes, and as platforms for reproducible artistic work for the wider experimental music research community.

As will be explained in [Section 4.3](#), it is within this context of creating, developing, implementing and iterating on a set of open design patterns for AR art and music that the research methods employed in the following chapters are carried out. This started with a motivation for creating smaller (in complexity and in interactable space) experiences, that were simple in their sensory modulation and instrumental reactivity, with the view to later incorporate them in larger scale interactive AR experiences once sufficiently tested. However, due to COVID-19 the prospect of user-testing and iterating on the design of multiple experiences, especially large scale ones, through participant feedback became increasingly difficult to facilitate and plan for. Thus, I opted instead to focus on developing working musical AR prototypes that I myself would play regularly, tweak, and iterate upon. Then, and there, an AR research and compositional practice began.

It's pertinent to mention that practical work, where carried out, has always been documented for analysis, open-research, and archival purposes. I found the combination of [my personal website](#) as well as [GitHub](#) a suitable solution for this. Links to project files will be included in the study chapters. Before outlining the aforementioned design patterns, I would first like to touch upon the concept of resistance, and its role in motivating the methodological approach taken in the thesis.

4.2 Resistance in Practice

Formative in the development of the methodological approach of this thesis has been the concept of ‘resistance’. In short, this could be characterised as a motivation for changing or acting against the state of the art, practice, technology, or philosophy, out of a discontent for its status quo. This isn’t to be confused by resistance in the sense of experiencing resistance to *do* something, although as might be expected, that has reared its head a few times over the last three years too. Its difficult to place resistance into the linear narrative of this thesis, some resistance formed my initial motivations for the carrying out of this research, and some was garnered along the way, especially around the time of the uptake in ‘Metaverse’ as a term to describe the sum total of all creative XR development. Either way, resistance has made up a large portion of my perspective on AR technologies, its role within digital music and humanities more broadly, and my rationale for pursuing several key areas discussed in the thesis so far (anti-consumerism, multisensory experience, agency), and to come.

As referred to in [Section 3.1](#), contemporary design research in the fields of computational art and music have in recent years rallied behind the broader “free/libre and open-source software” and “open-source hardware” ethoses (abbreviated to FLOSS and OSH respectively). This approach, often hand-in-hand with the Maker, DIY, and Hardware Hacking subcultures, stands in stark contrast to the goings-on of mega-corporations board rooms, with their design ‘sprints’, ‘agile’ workflows, and ‘human-centred’ design. While commentary on the considerable role of un(der)funded volunteer labour involved in OS tools and projects is outside of the scope of this thesis, it is worth mentioning. The approaches taken by these subcultures towards embodied design and composition workflows however, has been highly beneficial for the field. In the case of digital music making, software tools and coding languages like PureData, Maximilian, SuperCollider, ixilang, Sonic Pi, Tidal-Cycles, and Bela have granted students, researchers, and composers across disciplines, free and beginner-friendly access to a wealth of tutorials, examples, and tools for advanced real-time sampling and synthesis techniques.

It could be proposed, that this approach is motivated in part by a resistance against many of the ‘features’ typically found in commercial “closed-source” software and hardware. Providing significant barriers to accessible learning, creativity, and artistic authenticity, these include lack of interoperability with other software, use of proprietary file formats, high initial or recurring subscription costs, and restricted access to lower-level parameters,

functions, and settings. It is within this resistance against the closed-source and / or commercial that the practical work of this thesis attempts to situate itself, drawing on its implicit designerly practices and aesthetics. As a result, the path taken to address the key areas of the thesis has involved the bringing into the fore, knowledge from the creative practice of hacking, making, designing, and iterating with experimental and DIY hardware and software.

A second form of resistance that the methodological approach of this thesis explores is the ocularcentrism found in AR technologies mentioned previously in [Section 2.1.3](#) and [Section 3.3.2](#). As a practitioner of music, and audio related technologies, this has provided a significant portion of the motivation for carrying out the present research, as well as a significant portion of the challenges faced when attempting to develop AR as a medium for sound-based interactive experiences.

Although m

4.3 Design Patterns for AR Art and Music

The term design pattern here, is borrowed from the field of computer science, where it is taken to describe “communicating objects and classes that are customized to solve a general design problem in a particular context” ([Gamma, 1995](#)). A design pattern thus “names, abstracts, and identifies the key aspects of a common design structure that make it useful for creating a reusable object-oriented design”. So, while as a method it may not operate completely as it would in its native computer science, to address the outstanding aims of the thesis, design patterns do serve to be less rigid than a framework, more open to adaptation than guidelines; whilst inheriting the meaningful organisational structure that comes with an object-oriented design approach¹. Design patterns are characterised by having four elements:

- The **pattern name** describes the design problem at a higher level of abstraction
- The **problem** describes the situation in which you might apply the pattern
- The **solution** describes the relationships between elements of the pattern that aim to solve the problem
- The **consequences** are the results and trade-offs of applying the pattern

¹ explain OOD

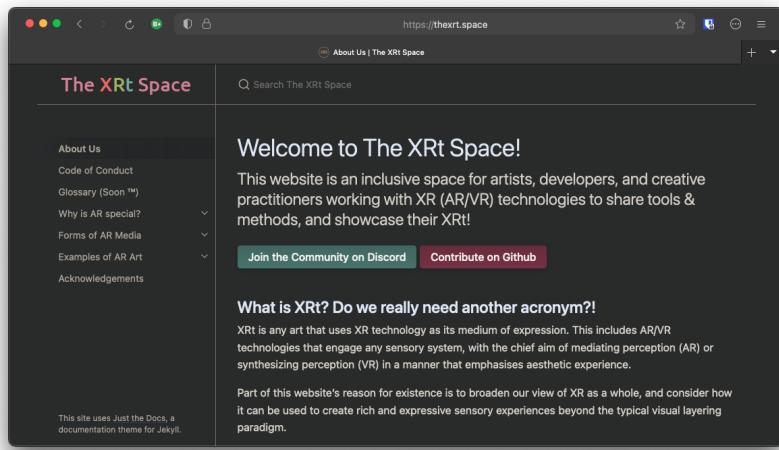


Figure 4.1: The XRt Space website

These design patterns are therefore subject to iteration, and the latest version can be found on [the XRt Space website](#), a community-editable repository created to host and update them. The principles used to guide the patterns draw on the resistances outlined in [Section 4.2](#), namely taking a DIY approach, decoupling from the ocularcentric and layering paradigms of typical AR experience, and attempting to navigate an inherently consumerist space whilst trying not to contribute to exploitative systems of oppression that uphold it. They are also guided by the theoretical proposals of [Chapter 3](#): that participant's and performer's cognitive processes in the experience of AR artworks are embodied, embedded, enacted, and extended, and have the potential to be modulated to extents that offer novel aesthetic experiences of augmented [materiality](#), [embodiment](#), and [space](#).

4.3.1 The Experience

[Chapter 3](#) drew on a number of theoretical propositions, and put forward that AR has the potential to scaffold new modes of performance and expression in the arts and music, furthermore that from an enactivist approach experience, this would consist in radically modulating the material, embodied, and spatial experience of participants.

The AR Subform

The Real-Virtual Dynamic

Sensory Engagement

The Interaction

4.3.2 The Instrument

Wearable

Tangible

Situated

4.3.3 The Environment

4.3.4 Balance between R/V

4.3.5 Allowance for the Real

4.4 AR Experience Studies

4.4.1 The area~ System - [Chapter 5](#)

The starting point for this was the creation of area~, a non-visual audio AR experience centred around the self-composition of a hybrid listening environment using an ambisonic microphone, head-, and hand-tracking. As will be expanded upon in [Chapter 5](#), due to the circumstances of the COVID-19 lockdown, I opted for an autobiographical design method for the iteration and evaluation of the system. Out of the evaluation of this project, namely the endeavour to expand into audiovisual AR experiences, I discovered a suitable platform to develop the rest of doctoral research on: the open-source Project North Star AR headset.

4.4.2 Evaluating polaris~ - [Chapter 6](#)

4.4.3 Experimental A/V AR Performance @ The Rosehill - [Chapter 7](#)

THE AREA~ SYSTEM

5

Exploring Real and Virtual Environments Through Gestural Ambisonics and Audio Augmented Reality

This chapter draws the majority of its content from a published article in the Goldsmith's journal Sonic Scope: New Approaches to Audiovisual Culture

(Bilbow, 2021a)

5.1 Computational 'ARt' with area~

This paper contributes towards the holistic development of AR applications by exploring one of the main attributes of an AR system: the combinational relationship of real and virtual environments, specifically within audio augmented reality (AAR). In doing so, I attempt to form a deeper understanding of AR's aesthetic and material capabilities as an artistic medium.

The area~ system is a gestural sound sampler that uses hand and head tracking to place and manipulate virtual audio sources in the user's environment, heard through bone conduction headphones which transmit sound directly to the cochlear without occluding the user's hearing. This allows the user to experience *virtual audio environments* overlaid seamlessly onto the *real audio environment*. Through gesture, the user can interact with and shape the combined *real and virtual audio environment* surrounding them.

This paper will address the research question: "*How can we better understand relationships between virtual and real environments through gesture and virtually placed AAR objects?*"

Understanding these relationships allows for the facilitation of engaging experiences within computational arts through richer multisensory human-computer interaction.

5.1.1 Technologies

The three technologies used in area~ are gestural hand tracking, rotational head tracking, and ambisonics. The gestural hand tracker used in the system is a Leap Motion LM-010 Controller, a USB infrared camera device that provides location and orientation data output of individual finger joints (and therefore hands) when they are presented above the device. The Leap Motion Controller (LMC) has been adopted in a multitude of settings such as being mounted on VR headsets ([Motion, 2016](#)), and converting hand gestures to MIDI (Leap Motion 2017). More recently, UltraLeap are investigating the use of this same technology with gesture-based public information screens to help combat the “hygiene risks of touch screens” ([2020b](#)).

Rotational head tracking is achieved via an inertial measurement unit (IMU). This small and inexpensive component provides orientational data output at 20 times a second. When affixed to the head via a headset or headphones, it is a relatively easy and cheap way of implementing head tracking into the system.

Ambisonics is an audio format that allows for full-spherical audio capture and playback ([Gerzon, 1973](#)), meaning that it includes sound sources above and below the listener as well as the conventional horizontal plane. There are four recorded channels (referred to as A-Format) that, unlike regular mono, stereo or surround sound formats, contain no information about the speakers that the signal should be delivered to. Rather, these channels can be encoded in such a way as to describe a three-dimensional field of sound referred to as B-Format, allowing the producer or artist to think in terms of sound sources in three dimensions rather than conventional speaker placement. B-Format can be decoded through “virtual microphones”, any number of which can be placed within this three-dimensional sound field to provide standard channel outputs.

For example, in area~, I have used a RØDE Soundfield NTSF-1 microphone array comprised of 4 microphones. The A-Format output is encoded to B-Format by an audio plugin. A software library decodes the B-Format to two responsive, binaural, virtual audio output channels. This all occurs in real-time, so that the microphones inside the three-dimensional sound field rotate proportionally as the user moves their head, providing realistic changes to what is heard.

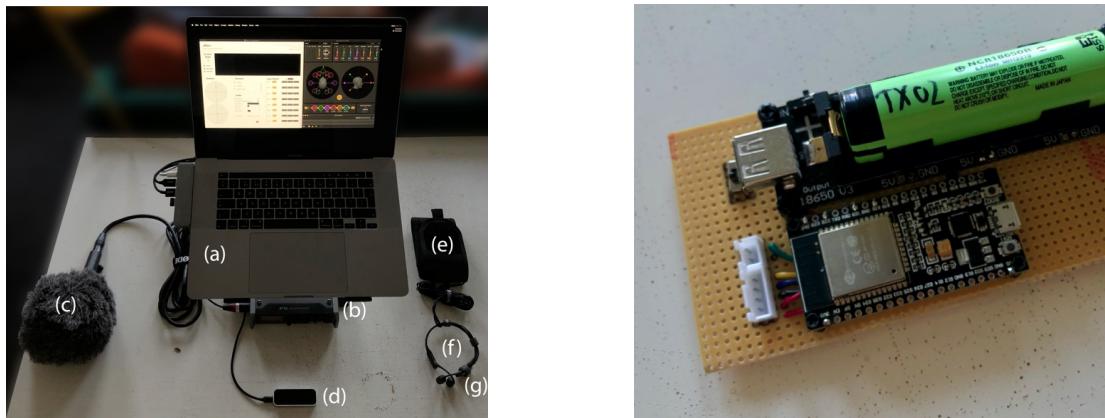


Figure 5.1: area~ hardware and PCB

5.2 The area~ system

The *area~* system, which stands loosely for ‘augmented reality environmental audio’ aims to afford users the ability to spectromorphologically (defined by Smalley to concern spatial, temporal and textural qualities of sound (1997)) manipulate sounds from their environment into a *virtual audio environment*. Through bone conduction headphones and head tracking, this sound field is heard in synchronicity with their actual environment. The system was created in order to explore and reveal the relationship between real and virtual environments.

5.2.1 Hardware Implementation

The on-desk hardware for the *area~* system shown in Figure 2 includes (a) a laptop running the Max MSP ([Cycling'74, 2020](#)) patch, (b) a 4 channel input audio interface, (c) an Ambisonic microphone, and (d) a Leap Motion Controller ([UltraLeap, 2020a](#)).

The wearable hardware used for the *area~* system comprises 2 sections: (e) a belt pouch containing a PCB-mounted ESP32 microcontroller ([Espressif, 2020](#)) and 18650 battery (shown in [Figure 5.1](#)), and (f) a pair of bone conduction headphones ([Aftershokz, 2020](#)), with (g) a mounted inertial measurement unit (IMU) for tracking head orientation.

The IMU and ESP32 are connected via a detachable 1.5m heat shrunk cable that runs from the back of the bone conduction headphones, down the length of the user’s back and into the belt-mounted transmitter pouch. With the integration of an Arduino library ([Winer, 2016](#)), the IMU data is transmitted to the laptop via Bluetooth from the ESP32. Audio is



Figure 5.2: area~ Max MSP patch

transmitted to the headphones via Bluetooth from the Max MSP patch running on the laptop.

The only hardware that needs to be accessible for the user is the Leap Motion Controller and the wearable hardware system. The laptop and audio interface are ideally hidden from the user. The microphone should be placed in a location that will provide the user with access to sounds that they wish to manipulate.

5.2.2 Software Implementation

The patch (Figure 5.2) uses the audio plugin (RØDE, 2020) shown in Figure 5.3 to encode the A-Format ambisonics microphone input into B-Format (a three-dimensional sound field), or what I will refer to as the *ambisonic palette*. This *ambisonic palette* is not heard by the user; instead, they can sculpt from it, forming their own audible (*B-Format*) *virtual audio environment* through hand gestures. I have defined these gestures in Max MSP with help from the IRCAM Leap Motion library (2014), and they occur over three stages of user interaction: **record, manipulate, spatialise**.

Record

The **recording or ‘sampling’ stage** is initiated by making a left-hand grab above the LMC, the longer lasting the grab, the longer the portion of audio from the *ambisonic palette* is sampled. The three-dimensional coordinates of the hand above the LMC correlates with the location of audio recorded (this is achieved by mapping the hand coordinates to a



Figure 5.3: RØDE Soundfield Plugin

virtual microphone inside the *ambisonic palette*), essentially allowing the user to record sounds around their person in three dimensions. Upon letting go of the grab gesture, the sample plays on repeat (using the karma~ Library (Constanzo, 2015)) through the bone conduction headphones, thus setting up the session's virtual audio environment.

Manipulate

The manipulation stage is automatically initiated after the ending of the previous grab gesture and uses translational (x, y, z) and rotational (roll, pitch) values from both hands when above the LMC. There are two audio effects being manipulated, with parameters from these effects mapped in different ways to the translation and rotation of the user's hands.

- The first effect is a band-pass filter which accentuates certain audio frequencies of the sample. The frequency, strength, and gain of the filter is determined by the parameter mappings detailed in [Figure 5.4](#)
- The second effect is a semi-random granular synthesiser. This selects and copies a section of the sample and deconstructs it into several hundred grains. The section of the sample granulated, and the individual grain duration is determined by the parameter mappings detailed in [Figure 5.5](#)

When the user decides to end manipulating the sample, they can do so by performing a grab with both hands. Once this happens, the band-pass filter and granular synthesis parameters are frozen for that sample.

Parameter	Mapping	Explanation
Centre Frequency	<ul style="list-style-type: none"> Point between both hands (x-axis) Linear map from LMC x-axis sensor range to BPF frequency (0 - 20kHz) 	Central frequency value (Hz) of filter.
Filter Strength	<ul style="list-style-type: none"> Absolute inter-hand distance (x-axis) Linear map from LMC x-axis sensor range to BPF Q value. 	Widen / narrow frequency range of filter
Gain	<ul style="list-style-type: none"> Average hand height (y-axis) Exponential map from LMC y-axis sensor range to 0.5 - 1.5x BPF gain 	Increase / decrease the volume of the filter

Figure 5.4: Band-pass filter parameter mappings

Parameter	Mapping	Explanation
Granulated sample section starting position	<ul style="list-style-type: none"> Right hand roll Linear map from LMC right hand roll sensor range (scaled 0 – 1) to 0 – 100% position on sample 	Where in this sample should I start the granulation?
Granulated sample section ending position	<ul style="list-style-type: none"> Right hand pitch Quadratic map from LMC right hand pitch sensor range (scaled 1 – 0 – 1) to 0 – 100% position on sample 	Where in this sample should I end the granulation?
Randomised grain duration lower bound	<ul style="list-style-type: none"> Left hand roll Linear map from LMC left hand roll sensor range (scaled 0 – 1) to grain duration lower bound (0ms – entire sample section length) 	Lower bound for randomised grain duration.
Randomised grain duration upper bound	<ul style="list-style-type: none"> Left hand pitch Quadratic map from LMC left hand pitch sensor range (scaled 1 – 0 – 1) to grain duration upper bound (0ms – entire sample section length) 	Upper bound for the randomised grain duration.

Figure 5.5: Granular synthesiser parameter mappings

Spatialise

The **spatialise stage** begins once the manipulation stage is ended by the user. The three-dimensional space above the LMC is mapped to the *virtual audio environment*, in which the user is currently listening to the sample that they have recorded. The user can use their right hand to move the sample around the *virtual audio environment*. For an example of the effect this has, moving the hand between the two extremes of the x-axis (left to right) results in hearing the sample move from ear to ear. The spatialise stage is ended by grabbing with the right hand.

Once the spatialise stage has ended, the user has the option to repeat the process 7 more times, allowing for the creation of a *virtual audio environment* comprised of up to 8 spatialised audio samples, or what I refer to as *nodes*.

Summary

The audio signal arriving in the two conduction pads in the headphones are the signals of two equidistantly spaced virtual microphones inside the *B-Format virtual audio environment* that decode it into two channels, left and right. Further interaction comes

from user head movement which, at all times, is mapped to the revolution of these two virtual microphones around the central point of the *virtual audio environment*. This means if there is a node playing to the left of the user, rotating the head 90° anticlockwise results in the *node* now sounding as if it is in front of the user's face. This is achieved via the ICST Ambisonics Library ([Schacher and Kocher, 2006](#)) and is elaborated on in the sections [Section 5.3.3](#) and [Section 5.4.1](#), but for now it is worth mentioning that it allows for immersion into a combined *real and virtual audio environment*.

To summarise, the patch can be categorised into having two inputs: audio from the user's environment and hand gesture, and one output: *the virtual audio environment*. In the background, this audio input is decoded into the *ambisonic palette* (inaudible), which is acted on by the user's hands to form one audible output: the *virtual audio environment*, which is comprised of up to 8 *nodes*. Through the choice of sensory overlay (bone conduction) and integration of head tracking, this *virtual audio environment* is experienced synchronously with the users' real, multisensory environment.

5.3 Study

5.3.1 Autobiographical Design

Originally, the study was planned for late March and would involve several user interaction studies. However, due to the UK lockdown in response to the unfolding COVID-19 pandemic, this was postponed until later in the year. Instead, I investigated the system using an autobiographical design method, framing it as a hypothesis study to better understand relationships between virtual and real environments, in the hope of developing a practice-led method for creating and researching multisensory augmented reality experiences.

Autobiographical design is defined as “design research drawing on extensive, genuine usage by those creating or building the system” ([Neustaedter and Sengers, 2012](#)). Neustaedter and Sengers define “genuine usage” here to mean that changes are “based on the true needs of the researchers, rather than them pretending to have needs expected of target users”.

Due to the lockdown, and therefore inability to conduct in-person user-tests, this research method was beneficial as I was spending large amounts of time with the system.

Moreover, there are several suggested requirements of employing this research method that Neustaedter and Sengers highlight that are true of the area~ system:

- The existence of a genuine usage of the system
- The system being already developed
- The ability for fast tinkering
- Record-keeping of the design process
- Long-term usage of the system

Furthermore, as AR technology moves towards being a component of future general personal computing, there is a need for first-person research methods that take into consideration the effects of prolonged system usage as well as the arising relationship between user and system (Desjardins and Ball, 2018). Moreover, these methods have been found to be specifically relevant to wearable systems (Cecchinato et al., 2017).

A disadvantage of using autobiographical design as a research method is its inability to establish generalisability (also the case with ethnography, case studies and participatory workshops), which is why I still intend to conduct wider usability studies in the future.

5.3.2 Design

The study was designed as a cycle, in order to promote fast tinkering, record-keeping, and long-term usage in line with autobiographical design guidelines.

- a) Over three sessions, ideally during the same week, the system is used with a logbook at hand in order to facilitate record-keeping of hardware setup, *node* manipulation and completed *real and virtual audio environment* listening experience remarks.
- b) After each session, the notes are formalised into a database and categorised as “user experience remarks” or “improvement remarks” (subject to increase in categories).
- c) At the end of the week, the three sessions’ notes are summarised into a “check-in” document, where user experience remarks are collated, and improvement remarks are further categorised into lists pertaining to the area of the system that needs improvement or change.
- d) Those changes are then made to the system and the cycle restarts.

5.3.3 Results

Hardware Location

I have observed that an environment with a lower noise floor is desirable and have implemented a normaliser to deal with re-recording loud background / ambient noises during successive *node* recording. I found myself basing the choice of microphone placement on what I wanted the *virtual audio environment* to sound like. Since the system uses an ambisonic microphone, consideration of the spherical 360° field of the microphone would lead to a richer *ambisonic palette* and subsequent *virtual audio environment*.

Two sessions were based outdoors and involved natural sounds such as birds, trees and wind, as well as passing cars. One of the sessions was inside and took place at the same time my partner was on a Zoom call, and therefore the *ambisonic palette* was invariably based on her speech and my movement and action inside the room. I want to look into placing the microphone inside bushes, trees, etc. rather than in open spaces to explore aesthetically the *virtual audio environment* that arises from such placement. Furthermore, the relationship between *real and virtual audio environment* would be quite different. Through their ears, the user would be rooted in their position in the environment, but because of the inherent blend between hearing and bone conduction, they would simultaneously experience the sonic environment of the bush mediated by the Max MSP patch.

Experiences

The system takes considerable time to set up, especially when documenting with video, audio, and notes. This process could certainly be streamlined further. Overall, I was pleased with the sound quality; the microphone picks up the environment very clearly. However, I remarked that the manipulation stage could feature more interesting real-time auditory effects on the *nodes*. The blending of *real and virtual audio environment* is achieved well via the bone conduction headphones and there was a subconscious registration via the head-tracking that gave me the very real impression that there was a 3D environment of *nodes* around my body.

The IMU on the bone conduction headphones sometimes provides erratic and erroneous data, leading to accidental revolutions of the *virtual audio environment* around the head. Despite technical difficulties such as this, in one autobiographical design session, when I

took the headphones off, I wrote that “I felt like I’d been disconnected from something” and that “my senses felt heightened before I took them off, not only to the *virtual audio environment*, but now more sensitive to the audio content of my real environment.

If I managed to capture an infrequent environment sound, such as a particular bird call, or a sentence of spoken word, the fact that the patch is set up to loop the samples gave a certain permanence to that otherwise impermanent sound. On multiple occasions I could not tell if the sounds of birds I was hearing originated from a *node* or from within a tree.

Arising Interactions

The maximum sample length is currently 28 seconds, but I have found myself mainly sticking to shorter loops, creating quite repetitive and rhythmic sequences. I remarked that grains from the synthesiser sounded like a permanent record of my gestures’ effect on the environment. I liked the playfulness of being able to record a sound from a certain location in my *real audio environment* and place it in different location in my *virtual audio environment*.

Despite the wearable hardware being wireless, hand interaction with the area~ system is inherently limited to being in range of the LMC (often placed on a table). I found myself wanting to be able to move around my environment whilst being able to record new nodes and hear existing nodes move relative to my body position.

Overall, the results from my autobiographical design method have shown that area~ is an effective tool for examining the combinatorial relationship between real and virtual environments. Despite the system’s hardware setup requiring some further work to allow for quicker start up and more accurate head tracking, it has provided me with novel aesthetic experience through:

- The blending of *real and virtual auditory environments* to create a third, augmented environment that was greater in experiential nature than the sum of its parts (not simply a combinatorial layering)
- The ability to spectromorphologically manipulate sounds in real-time in this third environment with the body
- The potential for creating believable illusions of real-world sound sources from these manipulated and spatialised virtual sounds.

5.4 Discussion

5.4.1 Audio Augmented Reality

In [Section 2.1.4](#), I detailed the importance of the relationship between real and virtual environments in the definition of AR systems. As a medium for creating computational art, the three specifications outlined by ([Azuma, 1997](#)) still hold true and should be discussed in the creation of such art or the development of their interactive systems of composition e.g. area~.

- a) Combines real and virtual
 - Using bone conduction headphones means that the user's unmediated perception of the environment is allowed to continue without interference, allowing the effective combination of real and virtual environments.
- b) Interactive in real-time
 - The system is made interactive in real-time through the gestural mappings (both head and hand) to sound effect parameters, and the perception of those effects through the mediating sensory display (bone conduction headphones).
- c) Registered in 3-D
 - The area~ system is registered in 3-D via the IMU data allowing for the mapping of head movement to *virtual audio environment*.
 - The mapping of hand movement (in the recording stage) to the spatial location recorded in the ambisonic palette (detailed in [Section 5.2.2](#)), also contributes to the notion that the *virtual audio environment* and real audio environment are aligned.

5.4.2 Interaction

The tables of parameter mappings shown in [Section 5.2.2](#) may seem like a chaotic mess, however, time has been spent making these mappings intuitive. For example, the representation of volume to a vertical scale corresponds with findings that non-musicians are relatively competent at attributing the size of air-gestures to heightened musical dynamics ([Godøy et al., 2006; Caramiaux et al., 2010](#)). As for the band-pass filter's horizontal pitch mappings, this is mainly based on the visual representation of frequencies on a horizontal

scale often found in the user interface of such filters. Nevertheless, it is pertinent to mention that field of psycho-musicology research (Timmers and Li, 2016) finds a correlation between pitch representation and horizontal space i.e. lower pitch to the left, higher pitch to the right. Although this is often attributed to the internalised representation of horizontal pitch on pianos by keyboard-playing musicians (Lidji et al., 2007; Rusconi et al., 2006), it has been found that this effect also propagates in non-musicians (Weis et al., 2016). This is also found to be the case in the “Playing air-instruments” study carried out by Godøy, Haga, and Jensenius (2006).

In contrast, interaction with the granular synthesiser is not made to be intuitive; instead, I have opted to hide or black-box the interaction through a mix of linear and quadratic mappings on each hand. This is in order to stir curiosity in the user and induce play, as found in [Section 5.3.3](#), but also to be tested in wider user studies, “I remarked that grains from the synthesiser sounded like a record of my gestures’ effect on the environment”.

Indeed, whilst outlining the material epistemologies of digital music instruments (DMIs), ([Magnusson, 2009b](#)) describes black-boxed DMIs as containing “knowledge of its inventors, which means that the users of the instrument do not need to have a deep knowledge of its internal functions”, furthermore clarifying that there is a “continued oscillation between a mode of conceptual (system design) engagement with the instrument and [an] embodied (performative) relationship with it”. This ‘oscillation’ displays, in turn, an underlying synergy between DMI development and the autobiographical design process, perhaps due to the similarities in requirements of the processes outlined in [Section 5.3.1](#). This synergy has led to the use of ABD in the development of many DMIs and interactive music systems ([Kiefer et al., 2020](#); [Martin, 2017](#); [Turchet, 2018](#); [Unander-Scharin et al., 2014](#)).

5.4.3 Computational Art

As a system for creating computational art in the form of in-situ AAR experiences, area~’s artistic output is firstly a real-time experience. In order to document these experiences of the system however, I have included the automated recording and saving of both the *ambisonic palette* (the ambisonic recording of the real audio environment), and the users *virtual audio environment* as separate B-Format .wav files in the project directory. These separate B-Format .wav files could be merged and decoded from B-Format to any number of speakers for a multi-channel installation of field recordings or manipulations of environments made with area~. This also leaves open the potential for area~ to be



Figure 5.6: area~ 360° video and ambisonic documentation (Bilbow, 2020)

used as a compositional tool, with potential applications in sound art and soundscape composition.

I also chose to merge and decode a set of these recordings to binaural ambisonic format, and have time-synced this recording with a 360° video and a screen recording of the patch that was taken during the system's use. This allows my performance of the virtual audio environment to be experienced second-hand with headphones. Dragging the screen in different directions with mouse/touchscreen emulates hearing the difference in environment if I moved my head in those directions. Optionally, an inexpensive smartphone VR headset (£5 - £15) can be used to heighten the interactivity of the experience. A screenshot of the 360° video (Bilbow, 2020) is shown in Figure 5.6. The potential for users' experiments with area~ to be captured and re-experienced interactively could have interesting applications, effectively allowing users to explore each others' lived aural experience of the system.

EVALUATING POLARIS~

6

An Audiovisual Augmented Reality Experience Build on Open-Source Hardware and Software

This chapter draws the majority of its content from a published conference paper in New Interfaces for Musical Expression

(Bilbow, 2022)

6.1 Design Framework

The polaris~ experience itself is built using mostly open-source hardware and software. As well as creating the experience, I was interested in keeping a log of its framework in order to ensure its reproducibility and to facilitate further creation of a wide variety of audiovisual AR experiences. Any ‘artist-developer’¹ wanting to work on similar experiences should have the ability, when following this framework, to rapidly create and prototype low-cost and privacy-respecting multisensory AR artworks, experiences, and instruments. This section details the framework, but additional information can be found on [Github](#) and [my website](#).



(a) Playing a virtual piano



(b) Resizing and moving virtual objects

Figure 6.1: Project North Star in use

6.1.1 polaris~ Hardware

Project North Star

The primary section of wearable hardware is the Project North Star (NS) AR headset, as open-sourced by LeapMotion (now Ultraleap) in 2018. It has a 3D-printable assembly, and its circuit boards, cables, and screens are available to buy [online](#); my North Star cost about £500 in total to build - 5 to 6 times less expensive than the commercial AR headsets mentioned, while maintaining industry-leading hand-tracking and 6DoF (6 degrees of freedom) depth tracking.

However, the time needed to build one and understand its workings well enough to troubleshoot any issues one faces may be a barrier to entry². Additionally, the finish material is not as polished as commercial headsets, and the overall size is larger and clunkier. It also needs to be tethered by USB and DisplayPort cable to a host computer or mobile compute pack.

Despite these drawbacks, my own experience of the headset has led to the rapid creation of many audiovisual AR prototypes. The fact that it requires no account to use, no developer's license to work with, and no data to be sent away to corporate servers has only added to my comfortability of using it as a creative tool.

¹ By using the term artist-developer, I refer to the media artist, creative coder, digital musician, or indeed any of the other many terms used to describe the category of artist who uses code and/or technology as their medium of artistic expression.

² I have found it a surprisingly integral part of the way I make sense of my creative practice, and fortunately there is a Project North Star Discord server with over 2400 helpful members that I have relied on heavily!



Figure 6.2: Project North Star worn with the addition of bone-conduction headphones

Bone Conduction

The secondary piece of hardware in the system is a pair of bone-conduction headphones, seen in [Figure 6.2](#). These have been used as a method of auditory display in several other AR projects ([Lindeman et al., 2008](#); [Barde et al., 2016](#); [Chevalier and Kiefer, 2018](#)), typically for their ability to deliver audio in an unobtrusive fashion, as well as their comfortability and cleanliness in installation settings. They do not obscure the wearer's hearing of their real environment, making them suitable for emphasising the intertwined nature of virtual and real components of an AR experience.

6.1.2 polaris~ Software

Unity

polaris~ uses an open-source software companion to the NS headset developed and maintained by Damien Rompapas ([2020](#)). At run-time, this Unity plugin (also making use of Ultraleap's Hands plugin) computes sensor readings from both the hand- and movement-tracking sensors and recreates the hand and headset pose in real-time inside the Unity scene. With the pose computed, it outputs the resultant view to the displays of the headset, rendering anything in the Unity scene relative to it.

Thanks to Unity's in-built audio spatialisation, any audio sample attached to an object in the Unity scene is, by default, spatialised in 3D and output via the bone-conduction headphones.

PureData

In my own research, the desire to implement more than just sample-based audio interactions led me to experiment with many different options for implementing real-time audio synthesis. I ranked each option I found by its ability to fulfil the below criteria: (a write-up can be found on [my website](#))

- a) Uses Unity's in-built audio spatialisation.
- b) Low computational cost on the host computer, and ability to be instantiated tens to hundreds of times procedurally.
- c) Ability to afford the artist-developer a wide palette of synthesis techniques.
- d) Allowing real-time parameter control of sounds via movement, gesture, and interaction with GameObjects in the Unity scene.
- e) Ability to rapidly prototype sound synthesis techniques and sonic interactions.
- f) Being free, open-source, and cross-platform.

Meeting all these requirements was the [LibPdIntegration](#) project developed and maintained by Niall Moody and Yann Seznec. It allows for the use of PureData patches in Unity, which, at run-time, are compiled to libPd (an embeddable version of PureData), and whose output is fed in real-time to the sound output of the GameObjects they are attached to ³.

To summarise, the use of the North Star companion software in conjunction with LibPdIntegration, allows for the creation of objects in an AR scene that can have their own PureData patches. These can be parameterised to gesture, movement, and interaction, in ways that manipulate their audio synthesis in real-time and at low computational cost. This all while existing three-dimensionally in the participants' visual and auditory fields. The fact that this is delivered via an optical-see through headset and bone-conduction headphones results in an experience that does not significantly hinder (compared to VR) the participants' ability to see and hear their real environment - augmenting reality rather than replacing it.

³ The inclusion of PureData means that the many artist-developers can rely on a program that has existed in the field for nearly 25 years, many of whom will have experience using it.

6.2 Study Design

In 2021, I ran a participant experience study of polaris~, the first experience I created with the above framework. The main objective of the study was to extract (via grounded theory) participant sentiment towards their ability to audiovisually express themselves through gesture and movement in the AR experience. Participants were recruited via university mailing-list, and were a mix of undergraduate and postgraduate Media, Film, and Music students.

6.2.1 Questionnaire

Participants completed a questionnaire, in which I asked for their age, gender identity, ethnicity, and occupation, to ensure a diverse and inclusive variety of participants. The mean age was ~24.7 years old, with a range between 19 and 37; there were 6 female and 4 male participants, belonging to a diverse group of ethnicities.

6.2.2 Tutorial

The participants were inducted into the experience via an introductory five-minute tutorial, the purpose being to ensure safety, build trust, and allow space for questions before beginning. This took the form of a narrated slideshow, in which I outlined the devices and interactions they should expect in the experience.

6.2.3 The polaris~ Experience

Once the participant was wearing the headset and headphones, confirmed they were comfortable and they were standing in the starting position ⁴, I began the Unity scene (a full demonstration of the scene can be found on my [YouTube channel](#)), and let them know that the experience was about to start.

The experience involved nine floating iridescent ‘orbs’, scattered at different distances from the starting position. The orbs would individually emit a repeating tone, whose pitch and tempo varied from orb to orb. Upon emitting a sound, the orb would eject a shower of white particles.

[iframe element]

⁴ As the headset was tethered to a computer, keeping the starting position the same between participants meant that the position of virtual content could be approximated to a clear space in the lab, marked by tape on the floor

Participants were invited to explore the space at their own pace. They could, of course, view the orbs from different angles, and hear their tones get louder and quieter as well as panning from ear to ear as they walked around them. I then prompted them to direct their gaze towards their hands, which were outlined, and when turning the left hand to face them, a menu appeared to the right of their palm. The menu contained two buttons, one labelled "Change Hand Colour", the other "Toggle Interactions". I prompted them to tap the top button, which upon depressing slightly and providing an auditory 'click', changed the colour of their hand's outline.

[iframe element]

Upon toggling the second button on, constant streams of particles started emitting from the centre of their palms. These particles persisted for approximately five seconds to conserve computational power. While the button was toggled on, and the streams were emitting, each hand also produced a continuous noise.

In addition to these interactions, a total of five further interactions existed in the experience. The first two concerned the position, orientation, and gesture of the hands, relative to the participant. Firstly, they could modulate the depth of a vibrato effect on the sound of the particle stream by gradually turning their palms towards their face. Secondly, they could affect the sound's filter cut-off frequency by pinching their fingers together into a point, resulting in a hissing sound; paired with the visual feedback of narrowing the particle stream.

[iframe element]

On pointing their palms in the direction of an orb, the particles would gravitate towards the orb, and begin to slow down, orbit, and rotate it. Depending on which palm was pointing towards the orb, there would be an additional effect that increased in intensity as the participant persisted in pointing towards the orb. When pointing their left palm, over the course of 20 seconds, the orb's tone and white particle burst increased in tempo. When pointing their right palm towards an orb, over the course of 5 seconds, the depth of a tremolo effect on that orb's tone increased, with the paths of the white particles emitted from the orb becoming more erratic and corkscrew-like.

[iframe element]

Exploration of the scene differed between participants, but once they had either found or been shown the interaction methods in the scene and had either explored their variety

or asked if there was anything else that they could do, I would ask them to try some experimental interactions.

Taking advantage of the vast number of parameters available to edit in Unity, I then changed elements of the visual experience ⁵ in real-time, asking for participant feedback on elements they preferred, and why. These elements of visual experience involved the orbs size, shape, gravity strength, and range; particle size, speed, lifetime, and colour.

[iframe element]

6.2.4 Interview

The next step of the study was a 10-minute interview, in which I asked participants about the positive and negative aspects of the experience, and anything they could suggest that would improve the experience. Questions were left deliberately broad due to the use of grounded theory to draw out themes for later analysis. I used a topic guide (see Appendix 1) to keep follow-on questions for replies related to the experience. This was in the form of a set of questions from the validated questionnaire titled "User Experience in Immersive Virtual Environments" by Katy Tcha-Tokey et al. (2016), which I adapted to suit AR.

6.3 Participant Feedback

6.3.1 Grounded Theory

I chose the constructivist strand of grounded theory as developed by Kathy Charmaz (2006) (building on the initial work by Glaser and Strauss (1967)) for my method of data analysis due to its ability to build theories from gathered data. Generally, generating a grounded theory extracts relevance from 'codes' - line-by-line summaries of transcribed speech, that are then iteratively and repeatedly refined, and sorted into categories. This contrasts approaching data collection with a specific hypothesis in mind.

In the constructivist version of the Grounded Theory method, rather than assuming emergence of 'unproblematic' and 'objective' categories from the data itself, categories are admitted to being 'mutually constructed' through interaction by the researcher with the data, considering and highlighting both the position and subjectivity of the researcher, as well as the partiality, and situational nature of the data itself.

⁵ Since the 11 PureData patches (9 for the orbs, and 1 each for the hands) were compiled at run-time they could not be edited on the fly

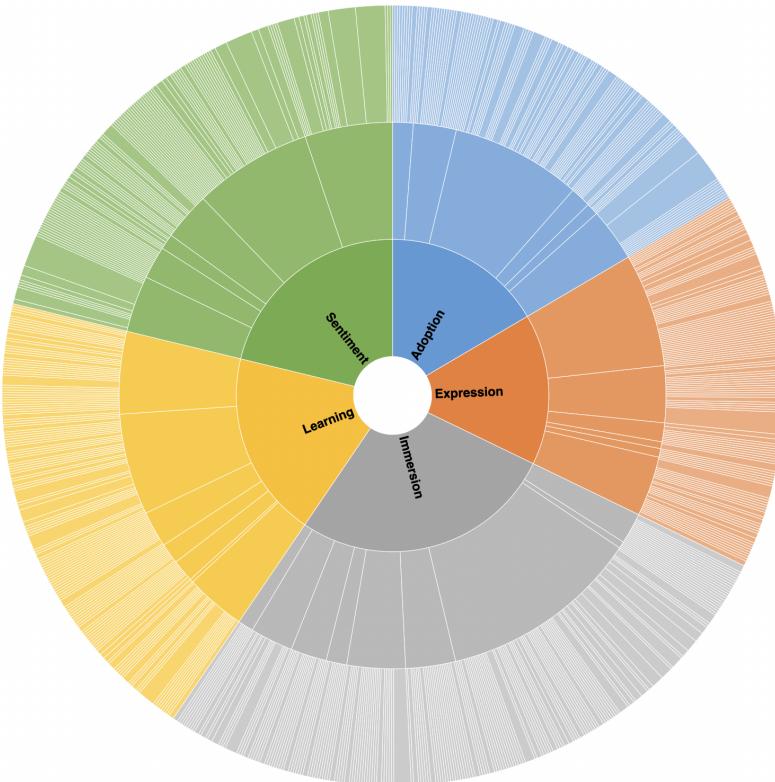


Figure 6.3: 5 categories, 37 subcategories and ~700 codes visualised in NVivo.

In this instance, a study where I set out to evaluate the experience of participants in an audiovisual AR scene, I believed that it was suited to provide a rich set of data from which to critique and iterate the experience, and eventually build a set of multisensory AR design guidelines.

Transcribing the nearly 7 hours of experiences and interviews, lead to 45,858 words, and approximately ~700 individual codes in the qualitative research software NVivo. These were sorted into 37 subcategories, with a total of 5 categories.

6.3.2 Sentiment

Included in this category are emotions elicited by the experience, the majority of which were related to novelty. Participants felt a mixture of emotions, most frequently wonder or awe at the visual components of the scene, and enjoyment of the scene components and the interactions with them.

One participant felt fixated by the experience: "*I am quite obsessed by this button*" (P9), whilst another felt satisfaction: "*Like a fascination, wonder, it was quite satisfying as well*" (P3).

All participants made ample use of simile and metaphor, likening visual elements to liquid, snow, fire, fireworks, magic, and confetti. The sound design was said to be "*ocean-like*" (P7) or "*wind-like*" (P1), and participants expressed similarities in their experience to 'being in' movies such as *Minority Report*, *Enter the Void*, and *Star Wars*.

6.3.3 Learning

From this category of codes, it is clear that the experience involved an element of learning different functions and abilities. One participant remarked: "*I was unsure, and I didn't know what would lead to a change in what I did, but after a point I understood*" (P6). Another linked learning as resulting in immersion: "*Each step that you learn something new about what you can do in that environment, that's when you become more and more immersed*" (P9).

6.3.4 Adoption

This category includes codes that referred to issues that could be a barrier to using the technology, recommendations for different utilisations of audiovisual AR, and codes relating to safety and accessibility implications of adopting AR.

Comfort and Fit

8 out of 10 participants struggled with the fit of the headgear on the Project North Star not being tight enough. As a result, some participants had to hold the headset with one hand throughout the experience. One participant expressed that this "*stood as a hurdle to [the] experience*" (P1), and another specified how this affected their involvement with experience in more detail: "*[Discomfort] doesn't ruin it, but it definitely brings me back to reality really quickly*" (P10).

Alignment and Tracking

Related to the above, were codes that referred to issues with alignment of content onto the real environment, such as the floating orbs and the outline of the hands. These were a product of the sensors, lighting, and content of the lab space. One participant even described the slight delay and misalignment of the hand outline as "*trippy*", going on to say that they "*sort of like that it's a bit out*" (P1).

Uses of AR and comparisons to other media

A wide variety of possible uses of audiovisual AR were highlighted by participants, including communication, conveying certain messages by highlighting important subject matter, art and music, virtual worlds, and video games. One participant, who studies Media for Development and Social Change and works with environmental NGOs, remarked that using audiovisual AR could help generate more interest in environmental conservation, because experiencing AR made them feel "*for some reason I act like this [virtual content] is more real than it is*", going on to say that "*rather than just watching it on the screen, you'd be more integrated*" (P6).

Another participant, studying on the same course, considered the use of AR in the documentation of the lived experiences of vulnerable people, such as refugees. They emphasised that compared to traditional media formats, AR was more "*interactive*", and that this could help in "*angling the participant to be in touch with the subject more*" (P1), possibly helping raise awareness of vulnerable people without tokenising their lived experience.

Several other participants described the potential uses of AR in artistic and musical contexts like polaris~. One participant said that AR had the potential to allow musicians to "*easily feel*" and "*play [...] with sounds*" (P2). One participant said that they could see AR being used for both instrument-building and creating installations (P3).

Safety and Accessibility

On the topic of safety, one participant expressed that despite "*[feeling] like there will be lots of benefits for [the uptake of AR and VR]*", they believed that it could lead people to "*lose track of reality*" (P7). Overall, participants reacted positively to the fact that their concerns over the comfortability and fit of the headset would be able to be addressed thanks to the open-source nature of the North Star's design.

6.3.5 Expression

Participants expressed themselves in various ways during the AR experience, and most described the ability to create visual and sonic components with their hands as the most compelling aspect of this expression. For example, one participant appreciated the variety of visual and sonic patterns that they could create: "*when your hands are together, it made one style of shape, and when your hands were away it would make different styles*

that affected the music" (P7), another emphasised variety as well as exploration: "*[The experience] let me explore by using my hands [and] doing different gestures*" (P6).

[iframe element]

Accordingly, several participants expressed their ability to act in the scene as "*control*" or "*power*". One participant, almost sounding guilty, said: "*I mean, it sounds weird, but I felt, like, very powerful*" (P10). Another, remarked similarly: "*I think once I realised that I had the power to add things, I didn't let go of that*" (P3).

While one participant remarked that "*physically doing something [that] you can see the effect of*" resulted in a feeling of "*control*" (P7), another participant noted that they didn't always feel this: rather, that this might grow over time and with increased familiarity of the scene, reinforcing the suggestion of a learning process in the experience (P3).

Closely related to these codes were ones where participants expressed their wish to have more power and control over the scene, or that they had wished for different outcomes. Most common was the wish to be able to move the orbs around themselves (P3, 6, 9) or the ability to change elements as I had done in the experimental part of experience themselves, and at their own leisure (P5, 7, 9, 10).

One participant agreed that adding further visual indicators for the effects being had on the sound would help them notice these changes. The same participant commented that the pinching gesture used to tighten the stream of particles from the hand was unintuitive and would have preferred it to be a pointing gesture (P2).

6.3.6 Immersion

Awareness

Most participants reported that they felt aware of their real-world surroundings during the experience, one commenting that it was "*because you could still see everything, you could still hear*" that they didn't get "*lost*" in the experience (P7). However, one of the participants was adamant that they had "*lost track of reality*" during the experience, and that it had felt "*like [they were] in another environment*" (P9), another still, observing that there were moments when they forgot where they were (P10). It's clear then, that the experience immersed the participants, some more than others. When asked to offer a rationale for their feelings of immersion, participants pointed to several factors.

Sights

One participant noted that they felt more immersed by the visual components of the scene, but that they would have felt less immersed if there wasn't an auditory component to the experience (P10). When being immersed by visual elements, for some it was colours that maintained this sense of immersion, with one participant commenting that the "*vividness*" (P8) and size of the colours when particles had been made larger in the experimental section is what led to the moments of highest immersion in their experience.

Sounds

On the topic of immersion through audio, one participant put this down to the feeling of being "*submerged*" in different layers of three-dimensional sound (P1). Similarly, many of the participants confirmed or independently reported to be able to discern and localise the tones they heard from different orbs around them, with some doing this by exploring through movement (P7), and others taking the visual cue of the white particle burst (P3). One participant remarked that the sound was the "*main aspect*" of immersion for them, and that it made them "*feel like part of*" the experience (P4); another commented that the sound "*surrounded*" them and held their concentration so much that they almost forgot that they were wearing the bone-conduction headphones (P2). For another still, it was the activation of one of the orbs' sound effects that made them feel part of the experience (P3).

Actions

A participant remarked that it was the feeling of "*creating*" that led to their immersion. Additionally, they remarked that it was the "*element of play*" in the interaction between particles and orbs that had immersed them in the experience; further noting that it was the way that the particles "*emerged*" from their hands that kept them engrossed (P9). On playfulness and fun, another participant commented that it was the fact that they could interact with content that wasn't "*there*" which was the most fun for them (P5).

Physicality of Content

One participant reported multiple times that they could "*feel*" the button when they pressed it (although technically it was floating adjacent to their hand in mid-air). They

agreed that it could be the mixture of the feedback sound, volumetric threshold trigger, and lighting that led to this effect (P9).

Another participant, upon placing their hands close to an orb remarked: "*I know there's nothing there, but when I put my hand [towards the orb] I feel like I'm going towards something warm*". This same participant expressed a keen interest during the experimental section in drawing large three-dimensional sculptures. During one of these drawings, the participant walked around their creation, and then took a moment before exclaiming: "*Oh! I forgot that it wasn't a real thing, I kept trying to go [around] it, but I could just go through it!*". Later, they noted: "*I guess it's about our brains. [...] When we see it in 3D we automatically think it's more real than [if it were on a screen]*" (P8).

6.4 Conclusion

Overall, the AR experience engaged participants fruitfully, with many noting their ability to express themselves audiovisually in creative ways. For polaris~, the ability to do so whilst maintaining a privacy-respecting and cost-effective focus, is a testament to the individual components of the framework used to create it, and the labour that has facilitated the development of these open-source solutions.

The categories of codes extracted from the grounded theory analysis have resulted in a rich set of data. From those related to participant sentiment, it's not only clear that the audiovisual AR experiences are able to elicit a wide variety of emotions, but that to explain and make sense of them participants often made use of metaphors and past experiences. The fact that it was expressed multiple times by participants that the experience was one that they'd never had before might be a way of accounting for this variety and overlap of emotions.

Relating to the category of learning, it's clear that participants sensed this from multiple aspects of the experience, with one pointing towards the dimensionality, and others pointing towards the interactive elements and the need to explore the scene. For one participant there was a connection between learning and immersion. Within the context of musical interfaces, this could be taken to show that viewing the participant as a learner in the experience could lead to a deeper level of engagement.

Within the category of adoption, the fact that the headset lacked a good fit for most participants was clearly what detracted from the experience most. It led not only to a reduction in experience immersion, but also to a lessening of comfort and the knock-on

effect of muscle fatigue and inability to exercise full agency for some participants. From the comments on different potential applications for AR, it is clear that the offer of deeper interaction with subject matter, increased immersion in an experience, and the sense of “*feeling*” or “*playing*” with virtual content has led to participants envisaging AR’s utilisation in several artistic and musical contexts. Notable was the suggestion that these facets of experience, especially that of the three-dimensional and context-specific sounds, could be employed to convey messages of socio-ethical importance as well as aesthetic experience. It is important to mention that some participants warned of negative side-effects or the potential for negative experiences in AR, showing the importance for a safety-minded approach to designing such AR experiences.

From the codes relating to self-expression and control, participants appreciated the ability to interact with their bare hands, and felt that this was the main contributor to their expression in the scene. This, for some, led to a feeling of power, and for some, a feeling of control over elements of the scene. For others, the feeling of control wasn’t entirely certain, with some noting that the scene had agency of its own. It is also conclusive from these codes that some participants desired more control over the scene, although implementing this would have to be done thoughtfully, without overloading the scene with content and parameters to change.

Immersion tended to stem from the fact that the elements of the experience: sights, sounds, and actions were spatialised in three dimensions. Participants enjoyed and felt immersed by the movement and colour of the visual elements and were able to discern and localise the source of different sounds in the experience; several noting that this element was the most immersive factor for them, whilst others preferred the visual elements. Others still, noted that the actions that they employed, gestural and movement-based are what immersed them in the experience most, and for some, this led to the virtual content of the scene feeling physical at times.

[iframe element]

6.4.1 Future Work

While the above analysis is not, on its own, sufficient to draw conclusive guidelines for design, it does shed light on several connections between elements of this specific experience: learning, comfort, safety, spatial perception, visual-, sonic-, and action-led immersion, and physicality of virtual content. In the future, I would like to focus on these

themes, and consider how they, as well as new developed experiences, might contribute to a set of multisensory AR design guidelines. Indeed, the framework used to create polaris~ has already been used in the creation of several experimental audiovisual AR performances.

[iframe element]

To improve on the comfortability of the experience, the first changes I made were to design an alternate headgear section for the headset. Over the course of a day, I collaborated with another NS community member on 3D printable designs that would allow the quick and easy substitution of the main headgear piece for a smaller one that was more accommodating of smaller head sizes.

I would also like to make the gestures and feedback for the orb-based sound effects more intuitive, reachable, and musical, as most participants relied on my showing them before they understood that it was a possibility in the scene. This could be done by conducting a trial with musician participants with which to iteratively develop these new interactions.

On the basis of the codes related to learning, I would like to experiment with an implementation of increasingly interactive and musical “*levels*”, both to reduce overwhelm when starting the scene, and to provide participants with a simple course through which to learn functions and express themselves accordingly.

6.5 Ethics Statement

The study followed University of Sussex ethics guidelines approved by the Social Sciences & Arts Cross-Schools Research Ethics Committee with the reference code: ER/SMB44/3.

6.5.1 Socio-economic Fairness

At all stages of this research, I have sought to minimise the outlay required to invest in technology needed to implement polaris~. This has meant opting for FLOSS (free / libre open-source software) where possible: The software companion to the North Star, Project Esky⁶, PureData, and OBS. Ideally the 3D-engine used would also be FLOSS, i.e. Godot, but due to time constraints I have opted for Unity which I am already experienced with, and which has great documentation and free learning materials.

⁶ ‘Esky’ is a colloquial term for an ice-cooler in Australia, etymologically deriving from the word ‘Eskimo’. I acknowledge the use of this term as being pejorative towards the indigenous populations of the Greenlandic and Canadian Inuit, Alaskan Iñupiat, and Yupik peoples; fully recognise their human and land rights, and have made the effort to redact the word within the main body of this article.

ough the Project North Star headset is OSH (open-source hardware), I have not had time to develop DIY and OSH audio solution yet, but any bluetooth bone-conduction headphones are compatible.

Unfortunately, the polaris~ experience (and as such, its framework) is not compatible with macOS and Linux, although it will be once Ultraleap provide their V5 drivers for hand-tracking on those operating systems.

It is worth noting that, as with most smaller suppliers of technology, the current global chip shortage has made getting hold of some components quite difficult at times, and recently the developers of the display driver board had to take the time to redesign it to get around this; the headset was not available to buy new for several months during this period. I suppose that this points to an issue: "*accessibility*" doesn't always correlate with "*availability*", especially when trying to circumvent expensive consumer technologies.

Since beginning these studies, Intel has discontinued their T261/T265 range of products, meaning that anyone looking to build the headset today would have difficulty including movement tracking, a fairly essential component of effective AR experiences. I am fully aware of the roadblock this puts in the way of polaris~ currently being reproducible as per its framework, but thankfully efforts are underway by developers in the community to solve this issue by migrating to Luxonis' modular and open-source range of tracking cameras; leaving the proprietary and closed-source hardware and software of Intel behind. If anything, their announcement has served as a reminder of the importance of FOSS/H components in community-led projects.

A note on the environment

While it is certainly difficult to mitigate the environmental impact entirely from projects that rely on technology, I have opted for compostable PLA material for the construction of the headset used in the study to reduce e-waste in the event of breakage or updates to the hardware design.

6.5.2 Study Participants

Accessibility

In anticipating the possibility of a diverse set of accessibility needs, I provided stepped and step-free access instructions to get to the lab, maintained clearly formatted corres-

pondence with participants via e-mail, and observed the university's COVID-19 policy during the study. Participants wearing glasses were able to use the headset with or without their glasses, and the volume of the AR scene was kept at a tolerable level, with the option given to participants to increase or decrease it.

Inclusion

Participants were recruited via internal University undergraduate and postgraduate mailing lists and were selected to ensure a diverse cohort of individuals from differing ethnicities, genders, and ages.

Remuneration

Study participants were compensated £15 each for the 45 minutes - 1 hour in which they took part in the study.

Consent

All participants provided written consent to take part in the questionnaire, experience, and interview; and also to being video and audio recorded during this time. They could at any time choose not to participate further, without being penalised or disadvantaged in any way. Their involvement in the study did not in any way impact their marks, assessments, or future studies. After analysis, their transcribed contributions were approved, and they had the opportunity to remove any anonymous, but still sensitive information.

Data and privacy

Data from the studies: video and audio recordings, transcriptions, and analysis files are secured on a separate hard drive and are password protected. All quotes and data have been anonymised in accordance with data protection legislation.

EXPERIMENTAL A/V AR PERFORMANCE @ THE ROSEHILL

7

quote

(Bilbow, 2022)

DISCUSSION**8***quote*

(Bilbow, 2022)

CONCLUSION

9

quote

(Bilbow, 2022)

**APPENDIX: 5 THE AREA~
SYSTEM****A**

**APPENDIX: 6 EVALUATING
POLARIS~**

B

**APPENDIX: 7 DEVELOPING
AN AR PERFORMANCE
PRACTICE**

C

BIBLIOGRAPHY

- Aftershokz, 2020. Aftershockz Aeropex.
- Armstrong, N., 2006. An Enactive Approach to Digital Musical Instrument Design. Ph.D. thesis, Princeton University.
- Azuma, R., 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* **6**:355–385.
- Barde, A., M. Ward, W. S. Helton, M. Billinghurst, and G. Lee, 2016. Attention Redirection Using Binaurally Spatialised Cues Delivered Over a Bone Conduction Headset. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **60**:1534–1538.
- Bilbow, S., 2020. Area~ 360° video / ambisonics documentation.
- Bilbow, S., 2021a. The area~ system: Exploring real and virtual environments through gestural ambisonics and audio augmented reality. *Sonic Scope: New Approaches to Audiovisual Culture* **2**.
- Bilbow, S., 2021b. Developing Multisensory Augmented Reality As A Medium For Computational Artists. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–7, Salzburg Austria. ACM.
- Bilbow, S., 2022. Evaluating polaris~ - An Audiovisual Augmented Reality Experience Built on Open-Source Hardware and Software. In *NIME 2022*, The University of Auckland, New Zealand. PubPub.
- Bimber, O. and R. Raskar, 2005. Spatial Augmented Reality: Merging Real and Virtual Worlds. A K Peters/CRC Press, zeroth edition.
- Brooks, J., S. Nagels, and P. Lopes, 2020. Trigeminal-based Temperature Illusions. In *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

Caramiaux, B., F. Bevilacqua, and N. Schnell, 2010. Towards a Gesture-Sound Cross-Modal Analysis. In *Gesture in Embodied Communication and Human-Computer Interaction*, volume 5934, pages 158–170. Springer Berlin Heidelberg, Berlin, Heidelberg.

Caudell, T. and D. Mizell, 1992. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, pages 659–669 vol.2, Kauai, HI, USA. IEEE.

Cecchinato, M. E., A. L. Cox, and J. Bird, 2017. Always On(line)?: User Experience of Smartwatches and their Role within Multi-Device Ecologies. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3557–3568, Denver Colorado USA. ACM.

Chalmers, D. J., 2022. Reality+: Virtual Worlds and the Problem of Philosophy. Allen Lane, an imprint of Penguin Books, London.

Charmaz, K., 2006. Constructing Grounded Theory. Sage Publications, London ; Thousand Oaks, Calif.

Chevalier, C. and C. Kiefer, 2018. Listening Mirrors: Prototyping for a Hybrid Audio Augmented Reality Installation. In José Alberto Gomes, Miguel Carvalhais, and Rui Penha, editors, *ICLI 2018, 4th International Conference on Live Interfaces. Inspiration, Performance, Emancipation.*, page 241, Porto, Portugal.

Chevalier, C. and C. Kiefer, 2020. What Does Augmented Reality Mean as a Medium of Expression for Computational Artists? *Leonardo* **53**:263–267.

Clark, A., 1999. An embodied cognitive science? *Trends in Cognitive Sciences* **3**:345–351.

Clark, A. and D. Chalmers, 1998. The Extended Mind. *Analysis* **58**:7–19.

Cohen, M., S. Aoki, and N. Koizumi, 1993. Augmented audio reality: Telepresence/VR hybrid acoustic environments. In *Proceedings of 1993 2nd IEEE International Workshop on Robot and Human Communication*, pages 361–364.

Constanzo, R., 2015. Tool: Karma~ (sampler/looper external) | Cycling '74.

Cycling'74, 2020. Max MSP.

Davies, C., 2004. Virtual space. In F. Penz, G. Radick, and R. Howell, editors, *Space: In Science, Art, and Society*, pages 69–104. Cambridge University Press.

- de Certeau, M., 1984. *The Practice of Everyday Life*. University of California Press, Berkeley, Calif.
- De Domenico, M. and H. Sayama, 2019. Complexity Explained.
- De Jesus, S. B., D. Austria, D. R. Marcelo, C. Ocampo, A. J. Tibudan, and T. Jhoselle, 2022. Play-to-Earn: A Qualitative Analysis of the Experiences and Challenges Faced By Axie Infinity Online Gamers Amidst the COVID-19 Pandemic. *International Journal of Psychology and Counseling* **12**.
- Delic, A. J. and P. H. Delfabbro, 2022. Profiling the Potential Risks and Benefits of Emerging “Play to Earn” Games: A Qualitative Analysis of Players’ Experiences with Axie Infinity. *International Journal of Mental Health and Addiction* .
- Desjardins, A. and A. Ball, 2018. Revealing Tensions in Autobiographical Design in HCI. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018 - DIS '18*, pages 753–764, Hong Kong, China. ACM Press.
- Dewey, J., 1934. Art as Experience. Capricorn Books, New York.
- Dey, A., M. Billinghurst, R. W. Lindeman, and J. E. Swan, 2018. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and AI* **5**:37.
- Di Scipio, A., 2003. 'Sound is the interface': From interactive to ecosystemic signal processing. *Organised Sound* **8**:269–277.
- Espressif, 2020. ESP32.
- Essl, G. and S. O’Modhrain, 2006. An enactive approach to the design of new tangible musical instruments. *Organised Sound* **11**:285–296.
- Fatemi, F., 2022. The Metaverse Will Radically Change Content Creation Forever.
- Feiner, S., B. MacIntyre, T. Höllerer, and A. Webster, 1997. A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. *Personal Technologies* **1**:8.
- Feiner, S., B. Macintyre, and D. Seligmann, 1993. Knowledge-based augmented reality. *Communications of the ACM* **36**:53–62.
- Frank, R., 2022. Metaverse real estate sales top \$500 million, and are projected to double this year.

Gach, E., 2022. Crypto Gaming 'Landlords' Upset They Can't Keep Exploiting All The Players Quitting.

Gallagher, S., 2017. Enactivist Interventions: Rethinking the Mind. Oxford University Press, Oxford New York, first edition edition.

Gamma, E., editor, 1995. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley Professional Computing Series. Addison-Wesley, Reading, Mass.

Gerzon, M., 1973. Periphony: With-Height Sound Reproduction. *Journal of The Audio Engineering Society* **21**:2–10.

Glaser, B. G. and A. L. Strauss, 1967. The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Transaction, New Brunswick.

Godøy, R. I., E. Haga, and A. R. Jensenius, 2006. Playing "Air Instruments": Mimicry of Sound-Producing Gestures by Novices and Experts. In S. Gibet, N. Courty, and J.-F. Kamp, editors, *Gesture in Human-Computer Interaction and Simulation*, volume 3881, pages 256–267. Springer Berlin Heidelberg, Berlin, Heidelberg.

Gottschalk, M., 2017. Move Over, Virtual Reality—a New Artistic Medium Is about to Emerge.

Grasset, R., E. Woods, and M. Billinghurst, 2008. Art and Mixed Reality: New Technology for Seamless Merging Between Virtual and Real. *IMedia-Space* page 10.

Günzel, S., 2019. What Do They Represent? Computer Games as Spatial Concepts. In E. Aarseth and S. Günzel, editors, *Ludotopia: Spaces, Places and Territories in Computer Games*, volume 63 of *Edition Medienwissenschaft*. transcript Verlag, Bielefeld, Germany, first edition.

Hayes, L., 2018. Live Electronic Music Performance: Embodied and Enactive Approaches. In *Proceedings of the 5th International Conference on Movement and Computing*, pages 1–3, Genoa Italy. ACM.

Hayes, L., 2019. Beyond Skill Acquisition: Improvisation, Interdisciplinarity, and Enactive Music Cognition. *Contemporary Music Review* **38**:446–462.

Hovhannisyan, G., A. Henson, and S. Sood, 2019. Enacting Virtual Reality: The Philosophy and Cognitive Science of Optimal Virtual Experience. In D. D. Schmorow and C. M. Fidopiastis, editors, *Augmented Cognition*, volume 11580, pages 225–255. Springer International Publishing, Cham.

- Hutmacher, F., 2019. Why Is There So Much More Research on Vision Than on Any Other Sensory Modality? *Frontiers in Psychology* **10**:2246.
- IRCAM, 2014. Leap Motion skeletal tracking in Max.
- Kane, I., 2022. Metaverse Meltdown: Top Metaverse Lands Lose 91% Value.
- Kiefer, C. and C. Chevalier, 2018. Towards new modes of collective musical expression through audio augmented reality. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 25–28, Virginia, USA.
- Kiefer, C., D. Overholt, and A. Eldridge, 2020. Shaping the behaviour of feedback instruments with complexity-controlled gain dynamics. In *NIME 2020*, page 6, Birmingham, UK. New Interfaces for Musical Expression.
- Leddy, T. and K. Puolakka, 2021. Dewey's Aesthetics.
- Ledesma, L., 2021. Axie Infinity Finds Ready Players in Hyperinflation-Racked Venezuela.
- Lefebvre, H., 1991. The Production of Space. Blackwell, Oxford, OX, UK ; Cambridge, Mass., USA.
- Lidji, P., R. Kolinsky, A. Lochy, and J. Morais, 2007. Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance* **33**:1189–1207.
- Lindeman, R. W. and H. Noma, 2007. A classification scheme for multi-sensory augmented reality. In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology - VRST '07*, page 175, Newport Beach, California. ACM Press.
- Lindeman, R. W., H. Noma, and P. G. de Barros, 2008. An Empirical Study of Hear-Through Augmented Reality: Using Bone Conduction to Deliver Spatialized Audio. In *2008 IEEE Virtual Reality Conference*, pages 35–42, Reno, NV, USA. IEEE.
- Lopes, P., S. You, A. Ion, and P. Baudisch, 2018. Adding Force Feedback to Mixed Reality Experiences and Games using Electrical Muscle Stimulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–13, Montreal QC, Canada. ACM Press.
- Maggioni, E., R. Cobden, and M. Obrist, 2019. OWidgets: A toolkit to enable smell-based experience design. *International Journal of Human-Computer Studies* **130**:248–260.

-
- Magnusson, T., 2009a. Epistemic Tools: The Phenomenology of Digital Musical Instruments. Ph.D. thesis, University of Sussex.
- Magnusson, T., 2009b. Of Epistemic Tools: Musical instruments as cognitive extensions. *Organised Sound* **14**:168–176.
- Mann, S., 1994. Mediated reality. Technical Report MIT-ML Percom TR-260, University of Toronto.
- Mann, S., 2018. Phenomenological Augmented Reality with the Sequential Wave Imprinting Machine (SWIM). In *2018 IEEE Games, Entertainment, Media Conference (GEM)*, pages 1–9, Galway. IEEE.
- Mann, S., J. C. Havens, J. Iorio, Y. Yuan, and T. Furness, 2018. All Reality: Values, taxonomy, and continuum, for Virtual, Augmented, eXtended/MiXed (X), Mediated (X,Y), and Multimediated Reality/Intelligence. In *Roceedings of the AWE 2018 Conference*, page 10, Santa Clara, CA, USA.
- Manovich, L., 2006. The poetics of augmented space. *Visual Communication* **5**:219–240.
- Marr, B., 2022. How To Buy Land & Real Estate In The Metaverse.
- Martin, C. P., 2017. Percussionist-Centred Design for Touchscreen Digital Musical Instruments. *Contemporary Music Review* **36**:64–85.
- Merleau-Ponty, M., 1945. Phenomenology of Perception. Routledge.
- Merleau-Ponty, M., 1968. The Visible and the Invisible: Followed by Working Notes. Northwestern University Press.
- Merrifield, A., 1993. Place and Space: A Lefebvrian Reconciliation. *Transactions of the Institute of British Geographers* **18**:516.
- Miéville, C., 1998. The conspiracy of architecture: Notes on a modern anxiety. *Historical Materialism* **2**:1–32.
- Milgram, P. and F. Kishino, 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information and Systems* **E77-D**:8.
- Mori, S., S. Ikeda, and H. Saito, 2017. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Transactions on Computer Vision and Applications* **9**:17.
- Motion, L., 2016. HTC Vive Setup.

- Narumi, T., S. Nishizaka, T. Kajinami, T. Tanikawa, and M. Hirose, 2011. Augmented reality flavors: Gustatory display based on edible marker and cross-modal interaction. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*, page 93, Vancouver, BC, Canada. ACM Press.
- Neustaedter, C. and P. Sengers, 2012. Autobiographical design in HCI research: Designing and learning through use-it-yourself. In *DIS '12: Proceedings of the Designing Interactive Systems Conference*, page 10, Newcastle Upon Tyne, United Kingdom.
- Nishida, J., S. Matsuda, H. Matsui, S.-Y. Teng, Z. Liu, K. Suzuki, and P. Lopes, 2020. HandMorph: A Passive Exoskeleton that Miniaturizes Grasp. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 565–578, Virtual Event USA. ACM.
- Nishida, J., S. Matsuda, M. Oki, H. Takatori, K. Sato, and K. Suzuki, 2019. Egocentric Smaller-person Experience through a Change in Visual Perspective. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Glasgow Scotland Uk. ACM.
- Noë, A., 2004. Action in Perception. Representation and Mind. MIT Press, Cambridge, Mass.
- Ongweso Jr., E., 2022. The Metaverse Has Bosses Too. Meet the ‘Managers’ of Axie Infinity.
- Papagiannis, H., 2014. Working towards defining an aesthetics of augmented reality: A medium in transition. *Convergence: The International Journal of Research into New Media Technologies* **20**:33–40.
- Papagiannis, H., 2017. The Critical Role of Artists in Advancing Augmented Reality. In *The Next Step: Exponential Life*, pages 124–139. Turner Libros.
- Puyat, M., 2021. Axie Infinity and the concept of play-to-earn. *CNN Philippines*.
- Riva, G., R. M. Baños, C. Botella, F. Mantovani, and A. Gaggioli, 2016. Transforming Experience: The Potential of Augmented Reality and Virtual Reality for Enhancing Personal and Clinical Change. *Frontiers in Psychiatry* **7**.
- RØDE, 2020. SoundField by RØDE Plugin.
- Rogers, A., 2021. Sci-Fi Icon Neal Stephenson Finally Takes on Global Warming. *Wired*.
- Rolland, J. P. and H. Fuchs, 2000. Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization. *Presence: Teleoperators and Virtual Environments* **9**:287–309.

Rompapas, D. C., D. F. Quiros, C. Rodda, B. C. Brown, N. B. Zerkin, and A. Cassinelli, 2020. Project Esky: Enabling High Fidelity Augmented Reality on an Open Source Platform. In *ISS '20: Proceedings of the 2020 ACM International Conference on Interactive Surfaces and Spaces*, Lisbon, Portugal.

Rosenberg, L., 1993. Virtual fixtures: Perceptual tools for telerobotic manipulation. In *Proceedings of IEEE Virtual Reality Annual International Symposium*, pages 76–82, Seattle, WA, USA. IEEE.

Rowlands, M., 2010. The New Science of the Mind: From Extended Mind to Embodied Phenomenology. MIT Press, Cambridge, Mass.

Rusconi, E., B. Kwan, B. Giordano, C. Umlita, and B. Butterworth, 2006. Spatial representation of pitch height: The SMARC effect. *Cognition* **99**:113–129.

Ryan, J., 1991. Some remarks on musical instrument design at STEIM. *Contemporary Music Review* **6**:3–17.

Schacher, J. and P. Kocher, 2006. Ambisonics Spatialization Tools for Max/MSP. In *Proceedings of the 2006 International Computer Music Conference*, volume 2006, pages 274–277, New Orleans, Louisiana, USA. Michigan Publishing.

Schiavio, A. and D. van der Schyff, 2018. 4E Music Pedagogy and the Principles of Self-Organization. *Behavioral Sciences* **8**:72.

Schraffenberger, H., 2018. Arguably Augmented Reality: Relationships between the Virtual and the Real. Ph.D. thesis, University of Leiden.

Schraffenberger, H. and E. van der Heide, 2015. Sonically Tangible Objects. In *xCoAx 2015: Proceedings of the Third Conference on Computation, Communication, Aesthetics and X*, pages 233–248, Glasgow, Scotland.

Schraffenberger, H. and E. van der Heide, 2016. Multimodal augmented reality: The norm rather than the exception. In *Proceedings of the 2016 Workshop on Multimodal Virtual and Augmented Reality - MVAR '16*, pages 1–6, Tokyo, Japan. ACM Press.

Seth, A. K., 2014. A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience* **5**:97–118.

- Seth, A. K., K. Suzuki, and H. D. Critchley, 2012. An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology* **2**.
- Skwarek, M., 2018. Augmented Reality Activism. In V. Geroimenko, editor, *Augmented Reality Art*, pages 3–40. Springer International Publishing, Cham.
- Sky Mavis, 2022. Axie Infinity: Gameplay.
- Slater, M., B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, 2010. First Person Experience of Body Transfer in Virtual Reality. *PLoS ONE* **5**:e10564.
- Slater, M., M. Usoh, and A. Steed, 1994. Depth of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* **3**:130–144.
- Smalley, D., 1997. Spectromorphology: Explaining sound-shapes. *Organised Sound* **2**:107–126.
- Stephenson, N., 1992. Snow Crash. A Bantam Spectra Book. Bantam Books, New York.
- Sutherland, I., 1968. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference*, pages 757–764.
- Suzuki, K., S. N. Garfinkel, H. D. Critchley, and A. K. Seth, 2013. Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia* **51**:2909–2917.
- Suzuki, K., W. Roseboom, D. J. Schwartzman, and A. K. Seth, 2017. A Deep-Dream Virtual Reality Platform for Studying Altered Perceptual Phenomenology. *Scientific Reports* **7**:15982.
- Suzuki, K., D. J. Schwartzman, R. Augusto, and A. K. Seth, 2019. Sensorimotor contingency modulates breakthrough of virtual 3D objects during a breaking continuous flash suppression paradigm. *Cognition* **187**:95–107.
- Tcha-Tokey, K., O. Christmann, E. Loup-Escande, and S. Richir, 2016. Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments. *International Journal of Virtual Reality* **16**:33–48.
- Thiel, T., 2018. Critical Interventions into Canonical Spaces: Augmented Reality at the 2011 Venice and Istanbul Biennials. In V. Geroimenko, editor, *Augmented Reality Art*, pages 41–72. Springer International Publishing, Cham.
- Thiel, T., C. Kozar, and I. Ünal, 2011. Invisible istanbul.

-
- Timmers, R. and S. Li, 2016. Representation of pitch in horizontal space and its dependence on musical and instrumental experience. *Psychomusicology: Music, Mind, and Brain* **26**:139–148.
- Turchet, L., 2018. Smart Mandolin: Autobiographical design, implementation, use cases, and lessons learned. In *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion - AM'18*, pages 1–7, Wrexham, United Kingdom. ACM Press.
- UltraLeap, 2020a. Leap Motion Controller.
- UltraLeap, 2020b. Touchscreens: Business as Usual Isn't an Option | Ultraleap.
- Unander-Scharin, C., A. Unander-Scharin, and K. Höök, 2014. The vocal chorder: Empowering opera singers with a large interactive instrument. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*, pages 1001–1010, Toronto, Ontario, Canada. ACM Press.
- Varela, F. J., E. Thompson, and E. Rosch, 1993. The Embodied Mind: Cognitive Science and Human Experience. MIT press, Cambridge (Mass.) London.
- Varese, E. and C. Wen-chung, 1966. The Liberation of Sound. *Perspectives of New Music* **5**:11.
- Veenhof, S. and M. Skwarek, 2010. We AR in MoMA.
- Veerapen, M., 2011. Encountering oneself and the other: A case study of identity formation in Second Life. In A. Peachey and M. Childs, editors, *Reinventing Ourselves: Contemporary Concepts of Identity in Virtual Worlds*, Springer Series in Immersive Environments. Springer London, London.
- Vermeulen, T., 2015. Space is the Place. *Frieze*.
- Vi, C. T., D. Ablart, E. Gatti, C. Velasco, and M. Obrist, 2017. Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition. *International Journal of Human-Computer Studies* **108**:1–14.
- Walther-Hansen, M. and M. Grimshaw-Aagaard, 2020. Don't extend! Reduce! The sound approach to reality. In *Proceedings of the 15th International Conference on Audio Mostly, AM '20*, pages 8–15, Graz, Austria. Association for Computing Machinery.
- Waters, S., 2007. Performance Ecosystems: Ecological approaches to musical interaction. In *Electroacoustic Music Studies Network EMS-07 Proceedings*, page 20, DeMonfort University, Leicester.

- Waterworth, J. and E. Waterworth, 2014. Altered, Expanded and Distributed Embodiment: The Three Stages of Interactive Presence, pages 32–45. DE GRUYTER OPEN.
- Weis, T., B. Estner, C. van Leeuwen, and T. Lachmann, 2016. SNARC (spatial-numerical association of response codes) meets SPARC (spatial-pitch association of response codes): Automaticity and interdependency in compatibility effects. *Quarterly Journal of Experimental Psychology* **69**:1366–1383.
- Willans, T., S. Rivers, and E. Prasolova-Førland, 2016. Enactive Emotion and Presence in Virtual Environments. In *Emotions, Technology, and Behaviors*, pages 181–210. Elsevier.
- Winer, K., 2016. ESP32 Arduino Library.
- Zavota, G., 2016. Expanding the Extended Mind: Merleau-Ponty's Late Ontology as Radical Enactive Cognition. *Essays in Philosophy* **17**:94–124.